

(19) Japanese Patent Office
(12) Official Gazette (A)
(11) Publication Number: Hei 11-3331
(43) Date of Publication: January 6, 1999
(51) Int. Cl. G06F 17/22

Request for Examination: Not yet submitted

Number of Claims: 13 (25 pages)

(21) Application Number: Hei 9-153627
(22) Date of Filing: June 11, 1997
(71) Applicant: Toshiba Corp.
[Translation of Address Omitted]
(72) Inventors: Yasushi ISHIZUKA et. al.
[Translation of Address Omitted]
(74) Representative: Patent Attorney Takehiko SUZUE et. al.

(54) [Title] Document Preparing Device, Predictive Inputting Method and Recording Medium

[Page 2 col. 1 line 1 to 24]

【Claim 1】 A document preparing device for preparing documents by performing kana-kanji character conversion on an inputted reading character string, comprising:

a dictionary for predictive input storing information for performing an input prediction based on a reading character string inputted for preparing a document;

a predictive processing means for acquiring a plurality of predictive character strings predicted when inputted, based on the inputted read character string for document preparation by referring to information stored in said dictionary for predictive input;

a dictionary for specifying a document field storing information based on the content of a prepared document;

a document field specifying means specifying the document field based on character strings in the text being prepared, by referring to information stored in the dictionary for specifying a document field;

a first narrowing means for narrowing a plurality of predictive character strings discriminated by the predictive processing means in

correspondence to the document field specified by the document field specifying means; and

a display means for displaying a predictive character string that has been narrowed by the first narrowing means,

wherein a predictive character string is inputted that has been selected from predictive character strings displayed by said display means.

[Page 7 col. 11 line 42 to 46]

【0051】 The dictionary 44 for specifying a document field stores information for specifying the field of the text currently being edited, for example the usage frequency of words according to their field, for input-predicting an optimal character string based on the reading character string input in accordance with the instructions of the user and the field of the text currently being edited.

* * *



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **11003331 A**

(43) Date of publication of application: 06.01.99

(51) Int. Cl.

G06F 17/22

(21) Application number: 09153627

(22) Date of filing: 11.06.97

(71) Applicant: **TOSHIBA CORP**

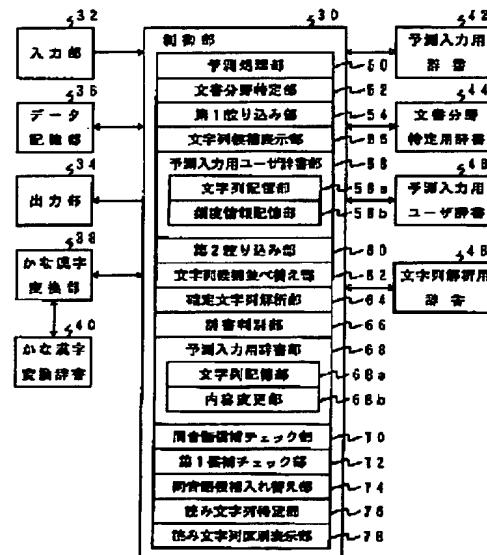
(72) Inventor: ISHIZUKA YASUSHI
NAKAZATO SHIGEMI
YANAKA SATORU

(54) DOCUMENT PREPARING DEVICE, PREDICTIVE INPUTTING METHOD AND RECORDING MEDIUM

(57) Abstract:

PROBLEM TO BE SOLVED: To reduce an operating burden by a user by securely predicting a character string the user desires to input.

SOLUTION: This device is provided with a dictionary for predictive input 42 registering information for inputting prediction based on a reading character string inputted for preparing a document, a predictive processing part 50 acquiring plural predictive character strings based on the read character string by referring to information registered in this dictionary 42, a dictionary 44 for specifying a document field registering information for specifying the field of a document based on the content of a preparing document, a document field specifying part 52 specifying the document field by referring to information registered in this dictionary 44, a first narrowing part 54 narrowing plural predictive character strings discriminated by the part 50 corresponding to the document field specified by the part 52, and a character string candidate display part 56 displaying a narrow predictive character string and inputs a predictive character string selected from the displayed predictive character strings.



COPYRIGHT: (C)1999,JPO

特開平11-3331

(43) 公開日 平成11年(1999)1月6日

(51) Int. C1. 6

G 06 F 17/22

識別記号

F I

G 06 F 15/20 5 2 0 S

審査請求 未請求 請求項の数 13 O L

(全 25 頁)

(21) 出願番号 特願平9-153627

(22) 出願日 平成9年(1997)6月11日

(71) 出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72) 発明者 石塚 靖

東京都青梅市末広町2丁目9番地 株式会社
東芝青梅工場内

(72) 発明者 中里 茂美

東京都青梅市末広町2丁目9番地 株式会社
東芝青梅工場内

(72) 発明者 谷中 悟

東京都青梅市末広町2丁目9番地 株式会社
東芝青梅工場内

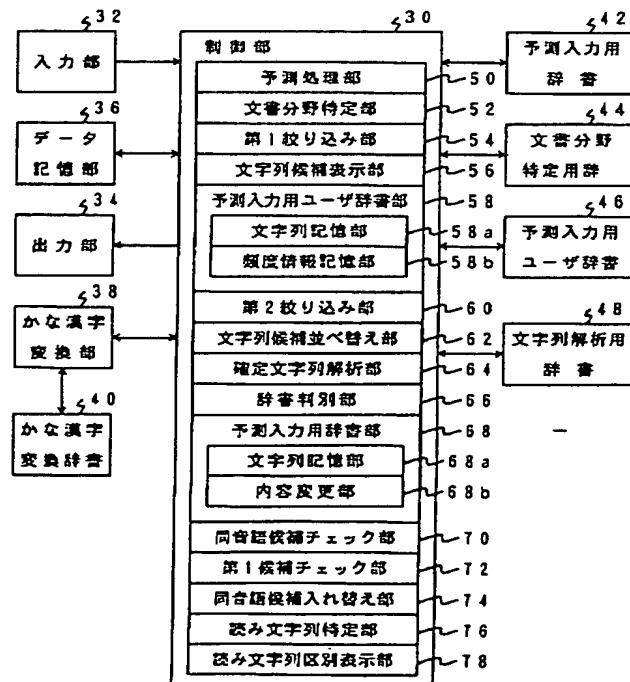
(74) 代理人 弁理士 鈴江 武彦 (外6名)

(54) 【発明の名称】文書作成装置、予測入力方法、及び記録媒体

(57) 【要約】

【課題】ユーザが入力しようと望んでいる文字列を的確に予測して、ユーザによる操作負担を軽減することを可能にする。

【解決手段】文書作成のために入力される読み文字列をもとにした入力予測を行なうための情報が登録された予測入力用辞書42と、この辞書42に登録された情報を参照して読み文字列をもとにして複数の予測文字列を取得する予測処理部50と、作成文書の内容をもとに文書の分野を特定するための情報が登録された文書分野特定用辞書44と、この辞書44に登録された情報を参照して文書分野を特定する文書分野特定部52と、予測処理部50によって判別された複数の予測文字列を、文書分野特定部52によって特定された文書分野に応じて絞り込む第1絞り込み部54と、絞り込まれた予測文字列を表示する文字列候補表示部56とを具備し、表示された予測文字列から選択された予測文字列を入力する。



【特許請求の範囲】

【請求項1】 入力された読み文字列についてかな漢字変換を行なうことで文書を作成する文書作成装置において、文書作成のために入力される読み文字列をもとにした入力予測を行なうための情報が登録された予測入力用辞書と、前記予測入力用辞書に登録された情報を参照して、文書作成のために入力される読み文字列をもとにして、入力されると予測される文字列である複数の予測文字列を取得する予測処理手段と、作成された文書の内容をもとにして文書の分野を特定するための情報が登録された文書分野特定用辞書と、前記文書分野特定用辞書に登録された情報を参照して、作成中の文書内の文字列をもとに、文書の分野を特定する文書分野特定手段と、前記予測処理手段によって判別された複数の予測文字列を、前記文書分野特定手段によって特定された文書の分野に応じて絞り込む第1絞り込み手段と、前記第1絞り込み手段によって絞り込まれた予測文字列を表示する表示手段とを具備し、前記表示手段によって表示された予測文字列から選択された予測文字列を入力することを特徴とする文書作成装置。

【請求項2】 前記文書分野特定用辞書には文書分野毎の各単語の使用頻度を示す情報が登録され、前記文書分野特定手段は、作成中の文書内の各文字列に対応する、前記文書分野特定用辞書に登録された単語の頻度情報が示す値の総和を各分野毎に求めて文書分野を特定することを特徴とする請求項1記載の文書作成装置。

【請求項3】 前記文書分野特定手段は、入力予測の対象とする文字列との相対的な位置関係に応じて、作成中の文書内の各文字列に対応する、前記文書分野特定用辞書に登録された単語の使用頻度を示す情報に重み付けを行なって総和を求めることを特徴とする請求項2記載の文書作成装置。

【請求項4】 前記予測入力用辞書には文書分野毎に各単語の使用有無を示すデータが設定された、分野毎使用可能状況データが登録されることを特徴とする請求項1記載の文書作成装置。

【請求項5】 入力すべきとして確定された文字列を登録するためのものであって、前記予測手段によって予測文字列を取得する際に前記予測入力用辞書と共に参照される予測入力用ユーザ辞書と、

予測文字列から選択されて入力された文字列を前記予測入力用ユーザ辞書に登録する文字列記憶手段と、前記予測処理手段によって取得された予測文字列が前記予測入力用ユーザ辞書に登録されていた場合に、この取得された文字列のみを予測文字列として絞り込む第2絞

り込み手段とを具備したことを特徴とする請求項1記載の文書作成装置。

【請求項6】 入力すべきとして確定された文字列と、この文字列に対応する使用頻度を示す頻度情報を登録するためのものであって、前記予測処理手段によって予測文字列を取得する際に前記予測入力用辞書と共に参照される予測入力用ユーザ辞書と、

入力すべきとして確定された文字列が、前記予測入力用ユーザ辞書に登録されていない場合に、この文字列を前記予測入力用ユーザ辞書に登録する文字列記憶手段と、入力すべきとして確定された文字列が、前記予測入力用ユーザ辞書に登録されていない場合に、前記文字列記憶手段によって登録された文字列と対応づけて頻度情報を登録し、前記予測入力用ユーザ辞書に登録されていない場合に、この文字列に対応する頻度情報を更新して登録する頻度情報記憶手段と、

前記予測処理手段によって前記予測入力用辞書と前記予測入力用ユーザ辞書に登録された情報を参照して得られた複数の予測文字列を、前記予測入力用ユーザ辞書に登録された文字列と前記頻度情報記憶手段によって登録された頻度情報に基づいて順番を並べ替える並べ替え手段とを具備したことを特徴とする請求項1記載の文書作成装置。

【請求項7】 文字列の解析を行なうための情報が登録された文字列解析用辞書と、

入力すべきとして確定された文字列について、前記文字列解析用辞書に登録された情報を参照しながら解析を行ない、前記予測入力用辞書に登録される単語を判別する確定文字列解析手段と、

前記確定文字列解析手段による解析によって得られた単語が前記予測入力用辞書に登録されているか否かを判別する辞書判別手段と、

前記辞書判別手段によって、前記確定文字列解析手段による解析によって得られた単語が前記予測入力用辞書に登録されていないと判別された場合に、この単語を前記予測入力用辞書に登録する文字列記憶手段とを具備したことを特徴とする請求項1記載の文書作成装置。

【請求項8】 前記予測入力用辞書には文書分野毎に各単語の使用有無を示すデータが設定された分野毎使用可能状況データが登録され、

前記辞書判別手段は、前記確定文字列解析手段による解析によって得られた単語が前記予測入力用辞書に登録されていると判別した際に、この単語について使用有無の何れを示すデータが設定されているか否かを判別するものであって、

前記辞書判別手段によって、前記確定文字列解析手段による解析によって得られた単語が前記予測入力用辞書に登録され、使用無しを示すデータが設定されていることが判別された場合に、使用有りを示すデータに変更する内容変更手段を具備したことを特徴とする請求項7記載

の文書作成装置。

【請求項9】 予測文字列を読み文字列として使用したかな漢字変換を行なうための情報が登録されたかな漢字変換辞書と、

読み文字列中に、予測文字列が含まれているかどうかを、前記予測入力用辞書に登録された情報を参照することによって判別する手段と、

この手段によって読み文字列中に予測文字列が含まれていると判別された場合に、前記かな漢字変換辞書に登録された情報を参照しながら、予測文字列を読み文字列として使用したかな漢字変換を行なうかな漢字変換手段とを具備したことを特徴とする請求項1記載の文書作成装置。

【請求項10】 読み文字列中に、選択された予測文字列の読みが含まれているかどうかを、前記予測入力用辞書に登録された情報を参照することによって判別する手段と、

この手段によって読み文字列中に前記予測文字列の読みが含まれていると判別された場合に、この予測文字列の読み文字列に対する文節区切りと同じ位置を優先した文節切りに応じたかな漢字変換を行ない複数の同音語候補を取得するかな漢字変換手段と、

かな漢字変換によって得られた複数の同音語候補中に、前記予測文字列に対応する読み文字列と同じ読み文字列に対応する同音語候補があるか否かを、前記予測入力用辞書を参照しながら判別する同音語候補チェック手段と、

前記同音語候補チェック手段によって同音語候補があると判別された場合に、この同音語候補が前記かな漢字変換手段によって得られた複数の同音語候補中の第1候補であるか否かを判別する第1候補チェック手段と、

前記第1候補チェック手段によって前記同音語候補チェック手段によって判別された同音語候補が第1候補でないと判別された場合に、この同音語候補の候補順位を第1候補に入れ替える同音語候補入替え手段とを具備したことを特徴とする請求項1記載の文書作成装置。

【請求項11】 入力されている読み文字列から予測入力に使用する読み文字列を、前記予測入力用辞書に登録された情報を参照して特定する読み文字列特定手段と、前記読み文字列特定手段によって特定された予測入力に使用する読み文字列と、その他の読み文字列とを区別して表示する読み文字列区別表示手段とを具備したことを特徴とする請求項1記載の文書作成装置。

【請求項12】 入力された読み文字列についてかな漢字変換を行なうことで文書を作成する際に、入力された読み文字列から入力予測を行なうための予測入力方法であって、

予測入力用辞書に登録された、文書作成のために入力される読み文字列をもとにした入力予測を行なうための情報を参照して、文書作成のために入力される読み文字列

をもとにして、入力されると予測される文字列である複数の予測文字列を取得し、

文書分野特定用辞書に登録された、文書の分野を特定するための情報を参照して、作成中の文書内の文字列をもとに文書の分野を特定し、

取得された複数の予測文字列を、特定された文書の分野に応じて絞り込み、

この絞り込まれた予測文字列を表示し、

この表示された予測文字列から選択された予測文字列を10 入力することを特徴とする予測入力方法。

【請求項13】 入力された読み文字列についてかな漢字変換を行なうことで文書を作成する際に、入力された読み文字列から入力予測を行なうためのプログラムであって、

予測入力用辞書に登録された、文書作成のために入力される読み文字列をもとにした入力予測を行なうための情報を参照して、文書作成のために入力される読み文字列をもとにして、入力されると予測される文字列である複数の予測文字列を取得し、

20 文書分野特定用辞書に登録された、文書の分野を特定するための情報を参照して、作成中の文書内の文字列をもとに文書の分野を特定し、

取得された複数の予測文字列を、特定された文書の分野に応じて絞り込み、

この絞り込まれた予測文字列を表示し、

この表示された予測文字列から選択された予測文字列を入力するようにコンピュータを制御するためのプログラムを格納したコンピュータ読取り可能な記録媒体。

【発明の詳細な説明】

30 【0001】

【発明の属する技術分野】 本発明は、入力された読み文字列についてかな漢字変換を行なうことで文書を作成する文書作成装置、予測入力方法、及び予測入力機能を実現するためのプログラムを記録した記録媒体に関する。

【0002】

【従来の技術】 一般に、日本語による文書を作成する文書作成装置では、入力された読み文字列についてかな漢字変換を行なうことで文字列を入力し、文書を作成している。

40 【0003】 かな漢字変換を行なうかな漢字変換機能は、ひらがなで入力された読み文字列を、かな漢字変換辞書に登録された情報を参照しながら漢字かな混じりの文字列に変換する。

【0004】 かな漢字変換機能は、ユーザによって入力された読み文字列に対応する漢字かな混じりの文字列を生成する。従って、ユーザは、作成しようとする文書のすべての読みを表す、ひらがなによる読み文字列を、キーボード等の入力装置を用いて正確に入力しなければならない。

50 【0005】 しかし、読み文字列を正確にすべて入力す

るという作業は、例えばキーボード操作に慣れていないユーザーにとっては多大な労力を必要とする。

【0006】そこで近年では、読み文字列を入力するための労力を少しでも軽減する機能として予測入力機能が利用され始めている。予測入力機能は、ユーザーが入力するであろうと思われる文字列を、入力しようとする文字列の全ての読みを入力することなく予測するというものである。

【0007】図25には、予測入力機能を用いた文字入力例を示している。例えば、図25(a)に示すように、「そふと」と入力した時点で、予測入力機能は、入力された文字列をもとにしてユーザーが入力しようとしている文字列を予測して、その入力候補となる文字列(予測文字列)を取得し、例えば図25(b)に示すように表示してユーザーに提示する。図25(b)に示す例では、入力文字列「そふと」に対して複数の予測文字列が取得されているため、予測文字列が一覧表示されている。

【0008】ここで、ユーザーは、提示された予測文字列から入力したい文字列を選択して、その文字列をそのまま確定入力することができる。さらに、図26に示すように、提示された予測文字列から選択した文字列に続けて読み文字を入力し、かな漢字変換を行うこともできる。

【0009】図26に示す例では、図26(a)のように入力された読み文字列「そふと」から予測された複数の予測文字列(図25(b))から、図26(b)に示すように「ソフトウェア」の文字列が選択された例である。この「ソフトウェア」の文字列に続けて、図26(c)に示すように「のひんしつ」の文字列が入力され、かな漢字変換の実行が指示されると、図26(d)に示すように、入力された読み文字列に対してかな漢字変換が行われ「ソフトウェアの品質」の漢字かな混じり文が生成される。

【0010】また、入力予測機能には、文字列の予測を行なう場合に、作成中の文書内の各単語の持つ点数の平均点から文書の意味ベクトルを算出し、その意味ベクトルとの距離が近い単語を予測文字列としてユーザーに提示する方法も考えられている(例えば、特開平7-271774号に記載されている)。

【0011】

【発明が解決しようとする課題】このようにして従来の文書作成装置では、入力効率を向上させるための方法として入力予測機能が用いられている。

【0012】しかしながら、前述した従来技術においては、入力された読み文字列をもとに予測された予測文字列が多い場合であっても、ユーザーに対してその予測して得られた予測文字列を全て提示している。従って、ユーザーは、提示された多数の予測文字列の中から希望する文字列を探し出し、選択する操作を行なわなければな

らなかつた。従って、多数の予測文字列が提示されるほど、文字列を選択するための処理に労力が費やされてしまうという問題があつた。

【0013】また、文書の意味ベクトルを使って予測を行う方法も、文書内の各単語の持つ点数の平均値を単純に求めて文書の意味ベクトルとして予測に使用しているため、この方法による文書の意味の特定精度は悪く、逆に間違った意味の文書と判断しユーザーが望んでいる文字列が予測されないといった問題が発生することも考えら
10 れている。

【0014】本発明は前記のような事情を考慮してなされたもので、ユーザーが入力しようと望んでいる文字列を的確に予測して、ユーザーによる操作負担を軽減することが可能な文書作成装置、予測入力方法、及び記録媒体を提供することを目的とする。

【0015】

【課題を解決するための手段】本発明は、入力された読み文字列についてかな漢字変換を行なうことで文書を作成する文書作成装置において、文書作成のために入力される読み文字列をもとにした入力予測を行なうための情報が登録された予測入力用辞書と、前記予測入力用辞書に登録された情報を参照して、文書作成のために入力される読み文字列をもとにして、入力されると予測される文字列である複数の予測文字列を取得する予測処理手段と、作成された文書の内容をもとに文書の分野を特定するための情報が登録された文書分野特定用辞書と、前記文書分野特定用辞書に登録された情報を参照して、作成中の文書内の文字列をもとに、文書の分野を特定する文書分野特定手段と、前記予測処理手段によって判別
20 された複数の予測文字列を、前記文書分野特定手段によって特定された文書の分野に応じて絞り込む第1絞り込み手段と、前記第1絞り込み手段によって絞り込まれた予測文字列を表示する表示手段とを具備し、前記表示手段によって表示された予測文字列から選択された予測文字列を入力することを特徴とする。

【0016】また、前記文書分野特定用辞書には文書分野毎の各単語の使用頻度を示す情報が登録され、前記文書分野特定手段は、作成中の文書内の各文字列に対応する、前記文書分野特定用辞書に登録された単語の頻度情報が示す値の総和を各分野毎に求めて文書分野を特定することを特徴とする。

【0017】また、前記文書分野特定手段は、入力予測の対象とする文字列との相対的な位置関係に応じて、作成中の文書内の各文字列に対応する、前記文書分野特定辞書に登録された単語の使用頻度を示す情報に重み付けを行なって総和を求める特徴とする。

【0018】また、前記予測入力用辞書には文書分野毎に各単語の使用有無を示すデータが設定された分野毎使用可能状況データが登録されることを特徴とする。

【0019】これにより、現在作成中の文書内の文字列

情報と文字列の位置情報を使って文書分野の特定を行なうことで文書分野特定の精度を高め、この方法で特定した文書分野を使って予測文字列候補を絞り込むことで、ユーザに対してはよく使われる最適な文字列だけを提示できるようになり、表示された予測文字列からユーザが選択する操作を簡単にすることができる。

【0020】また、入力すべきとして確定された文字列を登録するためのものであって、前記予測手段によって予測文字列を取得する際に前記予測入力用辞書と共に参照される予測入力用ユーザ辞書と、予測文字列から選択されて入力された文字列を前記予測入力用ユーザ辞書に登録する文字列記憶手段と、前記予測処理手段によって取得された予測文字列が前記予測入力用ユーザ辞書に登録されていた場合に、この取得された文字列のみを予測文字列として絞り込む第2絞り込み手段とを具備したことを特徴とする。

【0021】これにより、ユーザが以前に確定した文字列を使って予測文字列の候補を絞り込むことができるようなり、表示された予測文字列候補の中からユーザが選択する操作を簡単にすることができる。

【0022】また、入力すべきとして確定された文字列と、この文字列に対応する使用頻度を示す頻度情報を登録するためのものであって、前記予測処理手段によって予測文字列を取得する際に前記予測入力用辞書と共に参照される予測入力用ユーザ辞書と、入力すべきとして確定された文字列が、前記予測入力用ユーザ辞書に登録されていない場合に、この文字列を前記予測入力用ユーザ辞書に登録する文字列記憶手段と、入力すべきとして確定された文字列が、前記予測入力用ユーザ辞書に登録されていない場合に、前記文字列記憶手段によって登録された文字列と対応づけて頻度情報を登録し、前記予測入力用ユーザ辞書に登録されていない場合に、この文字列に対応する頻度情報を更新して登録する頻度情報記憶手段と、前記予測処理手段によって前記予測入力用辞書と前記予測入力用ユーザ辞書に登録された情報を参照して得られた複数の予測文字列を、前記予測入力用ユーザ辞書に登録された文字列と前記頻度情報記憶手段によって登録された頻度情報に基づいて順番を並べ替える並べ替え手段とを具備したことを特徴とする。

【0023】これにより、ユーザが以前に確定した文字列とその文字列の頻度情報を使って予測文字列候補の順番を並べ替えて表示することで、ユーザがよく使う文字列から表示されることになり、ユーザが選択する操作を簡単にすることができる。

【0024】また、文字列の解析を行なうための情報が登録された文字列解析用辞書と、入力すべきとして確定された文字列について、前記文字列解析用辞書に登録された情報を参照しながら解析を行ない、前記予測入力用辞書に登録される単語を判別する確定文字列解析手段と、前記確定文字列解析手段による解析によつ

て得られた単語が前記予測入力用辞書に登録されているか否かを判別する辞書判別手段と、前記辞書判別手段によって、前記確定文字列解析手段による解析によって得られた単語が前記予測入力用辞書に登録されていないと判別された場合に、この単語を前記予測入力用辞書に登録する文字列記憶手段とを具備したことを特徴とする。

【0025】これにより、確定した文字列が予測入力用辞書にない文字列である場合はその文字列を予測入力用辞書に追加することで、ユーザが今回確定した文字列を10次回の予測処理においてはこの文字列を予測文字列候補としてユーザに提示できるようになり、文書作成装置の予測入力環境を柔軟にユーザ毎の使用環境に合わせて変えていくことができるようになる。

【0026】また、前記予測入力用辞書には文書分野毎に各単語の使用有無を示すデータが設定された分野毎使用可能状況データが登録され、前記辞書判別手段は、前記確定文字列解析手段による解析によって得られた単語が前記予測入力用辞書に登録されていると判別した際に、この単語について使用有無の何れを示すデータが設定20されているか否かを判別するものであって、前記辞書判別手段によって、前記確定文字列解析手段による解析によって得られた単語が前記予測入力用辞書に登録され、使用無しを示すデータが設定されていることが判別された場合に、使用有りを示すデータに変更する内容変更手段を具備したことを特徴とする。

【0027】これにより、予測入力用辞書に記憶されている確定した文字列の状態データが、現在作成中の文書分野では使われない文字列であるというデータである場合は、そのデータを現在作成中の文書分野で使われる文字列であるというデータに更新することで、ユーザが今回確定した文字列を次回の同じ文書分野での予測処理においてはこの文字列を予測文字列候補としてユーザに提示できるようになり、文書作成装置の予測環境を柔軟にユーザ毎の使用環境に合わせて変えていくことができるようになる。

【0028】また、予測文字列を読み文字列として使用したかな漢字変換を行なうための情報が登録されたかな漢字変換辞書と、読み文字列中に、予測文字列が含まれているかどうかを、前記予測入力用辞書に登録された情報40を参照することによって判別する手段と、この手段によって読み文字列中に予測文字列が含まれていると判別された場合に、前記かな漢字変換辞書に登録された情報を参照しながら、予測文字列を読み文字列として使用したかな漢字変換を行なうかな漢字変換手段とを具備したことを特徴とする。

【0029】これにより、予測文字列をそのままかな漢字変換の読み文字列として使用できるようになり、ユーザが読み文字列中の予測文字列の区別がしやすくなる。また、予測文字列を曖昧性のない確定文字列情報と見なしかな漢字変換を行なうため、変換精度の向上も図るこ50

とができるようになる。

【0030】また、読み文字列中に、選択された予測文字列の読みが含まれているかどうかを、前記予測入力用辞書に登録された情報を参照することによって判別する手段と、この手段によって読み文字列中に前記予測文字列の読みが含まれていると判別された場合に、この予測文字列の読み文字列に対する文節区切りと同じ位置を優先した文節切りに応じたかな漢字変換を行ない複数の同音語候補を取得するかな漢字変換手段と、かな漢字変換によって得られた複数の同音語候補中に、前記予測文字列に対応する読み文字列と同じ読み文字列に対応する同音語候補があるか否かを、前記予測入力用辞書を参照しながら判別する同音語候補チェック手段と、前記同音語候補チェック手段によって同音語候補があると判別された場合に、この同音語候補が前記かな漢字変換手段によって得られた複数の同音語候補中の第1候補であるか否かを判別する第1候補チェック手段と、前記第1候補チェック手段によって前記同音語候補チェック手段によって判別された同音語候補が第1候補でないと判別された場合に、この同音語候補の候補順位を第1候補に入れ替える同音語候補入替え手段とを具備したことを特徴とする。

【0031】これにより、予測文字列を含む読み文字列をかな漢字変換した場合、予測文字列が変換結果である同音語候補の第1候補として表示されることで、ユーザはかな漢字変換後も予測文字列を候補選択の操作なしに常に取得できる。

【0032】また、入力されている読み文字列から予測入力に使用する読み文字列を、前記予測入力用辞書に登録された情報を参照して特定する読み文字列特定手段と、前記読み文字列特定手段によって特定された予測入力に使用する読み文字列と、その他の読み文字列とを区別して表示する読み文字列区別表示手段とを具備したことを特徴とする。

【0033】これにより、ユーザが読み文字列を入力中に予測入力実行指示を行なう場合、どの読み文字列に対して予測入力が行われるか分かるようになる。

【0034】

【発明の実施の形態】以下、図面を参照して本発明の実施の形態について説明する。図1は本実施形態に係わる文書作成装置のシステム構成を示すブロック図である。本実施形態における文書作成装置は、例えば磁気ディスク等の記録媒体に記録されたプログラムを読み込み、このプログラムによって動作が制御されるコンピュータによって実現される。

【0035】図1に示すように本実施形態における文書作成装置は、CPU10、メモリ12、表示装置14、入力装置16、出力装置18、及び外部記憶装置20によって構成されている。

【0036】CPU10は、文書作成装置全体の制御を

司るもので、各種のハードウェア装置とバスを介して接続されており、各装置の制御、装置間のデータの転送などの処理を行なう。CPU10は、メモリ12に格納されたプログラムに従って動作することで予測入力機能、かな漢字変換機能を含む日本語処理機能を実現する。

【0037】メモリ12は、プログラム領域12a、データ記憶領域12bが設けられ、CPU10の動作を制御するためのプログラムや各種データが必要に応じて記憶される。日本語処理システムを利用する場合には、プログラム領域12aに日本語処理プログラム24が格納される。日本語処理プログラム24には、予測入力機能を実現するための予測入力機能プログラム24a、かな漢字変換機能を実現するためのかな漢字変換プログラム24bが含まれる。データ記憶領域12bには、日本語処理システムによるかな漢字変換によって作成された文書の文書データ等が格納される。

【0038】表示装置14は、カラー液晶ディスプレイ及びそのコントローラから構成されており、CPU10の制御のもとで、日本語処理の処理の実行に伴う、かな漢字変換により文書を作成するために入力された読み文字列、かな漢字変換結果（漢字かな混じりの文書）、予測入力機能により予測された文字列（予測文字列）等の表示を行なう。

【0039】入力装置16は、キーボードやマウス等のポインティングデバイスからなり、かな漢字変換による文書作成のための読み文字列を含む各種のデータ及び命令を入力してCPU10に出力する。

【0040】出力装置18は、紙媒体に印刷を行なうためのプリンタ装置や、外部とのデータ通信を行なうための装置等である。

【0041】外部記憶装置20は、ハードディスク等の記憶媒体及びコントローラからなり、表示対象となる文書や、日本語処理を行なう際に参照される各種辞書のデータが納される。辞書には、かな漢字変換辞書20a、予測入力用辞書20b、文書分野特定用辞書20c、予測入力用ユーザ辞書20d、文字列解析用辞書20eが含まれている。辞書のデータは、必要に応じて読み出されてメモリ12に格納される。各辞書に登録されるデータの詳細については後述する。

【0042】次に、図1に示すシステム構成によって実現される本実施形態における文書作成装置の機能構成について説明する。図2は、本実施形態における文書作成装置の機能構成を示すブロック図である。

【0043】図2に示すように文書作成装置は、制御部30、入力部32、出力部34、データ記憶部36、及びかな漢字変換部38が設けられている。また、日本語文の文書を作成する際に用いられる、かな漢字変換辞書40、予測入力用辞書42、文書分野特定用辞書44、予測入力用ユーザ辞書46、文字列解析用辞書48の各辞書が用意されている。

【0044】制御部30は、CPU10がプログラムに従って動作することで実現されるもので、日本語処理システムの中心的な機能を果たし、入力装置16に対する操作によって入力部32から入力されたデータ等に応じて各機能部を制御する。予測入力機能の処理全般も制御部30が行なう。

【0045】入力部32は、入力装置16に対する操作に応じてデータやファンクション実行指示を入力して制御部30に渡す。

【0046】出力部34は、制御部30の制御のもとで、表示装置14におけるユーザが入力した読み文字列、かな漢字変換結果、予測入力機能によって予測された予測文字列等の表示や、出力部34からのデータ出力等を制御する。

【0047】データ記憶部36は、制御部30による各種処理に用いられるデータを記憶するもので、例えばかな漢字変換処理に必要なデータや草説明中の文書の文書(文字列)データ、文字列データを使った分野特定用のデータなど予測入力機能に必要なデータを記憶する。

【0048】かな漢字変換部38は、制御部30の制御のもとで、入力部32から入力された文書作成のための読み文字列のデータをもとに、かな漢字変換辞書40に登録された情報(単語情報、共起情報等)を参照して、入力文字列をかな漢字混じり文による文書に変換する。また、かな漢字変換部38は、予測入力機能を用いて予測され、文書作成のために選択された予測文字列を、かな漢字変換の対象となる読み文字列として使用する、かな漢字変換を行なうことができる。

【0049】かな漢字変換辞書40は、かな漢字変換部38によるかな漢字変換処理に必要な単語情報や、単語間の共起関係を示す共起情報等が登録されている。かな漢字変換辞書40に登録された共起情報は、予測文字列を読み文字列として使用するかな漢字変換を行なう際にも参照される。

【0050】予測入力用辞書42は、入力部32から入力された読み文字列をもとにして、予測入力機能によって文字列を予測するための情報が登録されている。予測入力用辞書42には、予測文字列の候補として提示される文字列が読みと対応づけて登録され、各文字列の各文書分野毎の使用可能状況が点数で表わされたデータ構造となっている。

【0051】文書分野特定用辞書44は、ユーザの指示により入力された読み文字列や作成中の文書の分野に基づいた最適な文字列を入力予測するため、作成中の文書の分野を特定するための情報、例えば文書分野毎の各単語の使用頻度を示す情報が登録されている。

【0052】予測入力用ユーザ辞書46は、入力された読み文字列についてかな漢字変換を行なうことで入力された文字列、すなわちユーザによって入力すべきとして確定された文字列が登録されるもので、予測入力機能に

よって入力文字列を予測する際に予測入力用辞書42と共に参照される。

【0053】文字列解析用辞書48は、入力された読み文字列についてかな漢字変換を行なうことで入力された文字列、すなわちユーザが入力すべきとして確定した文字列に対して、予測入力用ユーザ辞書46に登録する文字列を取得するための解析を行なうための情報が登録される。文字列に対する解析としては、形態素解析、構文解析などを行なうものとし、これらの解析用の情報が登録されるものとする。

【0054】また、図2に示す制御部30には、予測処理部50、文書分野特定部52、第1絞り込み部54、文字列候補表示部56、予測入力用ユーザ辞書処理部58、第2絞り込み部60、文字列候補並べ替え部62、確定文字列解析部64、辞書判別部66、予測入力用辞書処理部68、同音語候補チェック部70、第1候補チェック部72、同音語候補入替え部74、読み文字列特定部76、及び読み文字列区別表示部78の各機能部が設けられている。

【0055】予測処理部50は、予測入力用辞書42及び予測入力用ユーザ辞書46に登録された情報を参照して、文書作成のために入力される読み文字列をもとにして、ユーザによって入力されると予測される文字列である複数の予測文字列を取得する処理を実行する。

【0056】文書分野特定部52は、文書分野特定用辞書44に登録された情報を参照して、作成中の文書内の文字列と、文字列の位置情報を用いて文書分野を特定する処理を行なう。

【0057】第1絞り込み部54は、予測処理部50によって判別された複数の予測文字列を、文書分野特定部52によって特定された文書の分野に応じて絞り込む処理を行なう。

【0058】文字列候補表示部56は、予測入力機能によって予測された予測文字列を表示するものであって、第1絞り込み部54によって絞り込まれた予測文字列、文字列候補並べ替え部62によって並べ替えが行われた予測文字列の表示を行なう。

【0059】予測入力用ユーザ辞書処理部58は、予測入力機能による処理に伴う予測入力用ユーザ辞書46に対する処理を行なうもので、文字列記憶部58aと頻度情報記憶部58bを含んでいる。文字列記憶部58aは、入力された読み文字列についてかな漢字変換を行なうことで入力された文字列が、予測入力用ユーザ辞書46に登録されていない場合に、この文字列を予測入力用ユーザ辞書46に登録する処理を行なう。頻度情報記憶部58bは、入力された読み文字列についてかな漢字変換を行なうことで入力された文字列が、予測入力用ユーザ辞書46に登録されていない場合に、文字列記憶部58aによって登録された文字列と対応づけて頻度情報を登録し、入力された文字列が予測入力用ユーザ辞書46

に登録されている場合に、この文字列に対応する頻度情報報を更新して登録する処理を行なう。

【0060】第2絞り込み部60は、入力された読み文字列を使って予測処理部50によって取得された予測文字列が予測入力用ユーザ辞書46に登録されていた場合に、この取得された文字列のみを予測文字列として絞り込む処理を行なう。

【0061】文字列候補並べ替え部62は、予測処理部50によって予測入力用辞書42と予測入力用ユーザ辞書46に登録された情報を参照して得られた複数の予測文字列の順番を、予測入力用ユーザ辞書46に登録された文字列と頻度情報記憶部58bによって登録された頻度情報に基づいて並べ替える処理を行なう。

【0062】確定文字列解析部64は、入力された読み文字列についてかな漢字変換を行ない、入力すべきとして確定された文字列について、文字列解析用辞書48に登録された情報を参照して解析する処理、すなわち前記予測入力用辞書に登録される単語（例えば自立語）を判別するための処理を行なう。

【0063】辞書判別部66は、入力された読み文字列についてかな漢字変換を行なうなどして、入力すべきとして確定された文字列が、予測入力用辞書42に登録されているか否かを判別する処理を行なう。

【0064】予測入力用辞書処理部68は、予測入力機能による処理に伴う予測入力用辞書42に対する処理を行なうもので、文字列記憶部68a、内容変更部68bが設けられている。文字列記憶部68aは、確定された文字列が予測入力用辞書42に登録されていない場合に、その文字列を予測入力用辞書42に追加する処理を行なう。内容変更部68bは、確定された文字列が予測入力用辞書42にあるが、確定された時点での文書分野ではその確定した文字列が使われない文字列であるという状態である時に、その確定された文字列の状態を現在作成中の文書分野でよく使う状態に予測入力用辞書42の内容を変更する処理を行なう。

【0065】同音語候補チェック部70は、かな漢字変換によって得られた同音語候補中に、予測入力機能によって予測された予測文字列に対応する読み文字列と同じ読み文字列に対応する同音語候補があるか否かを、予測入力用辞書42を参照しながら判別する処理を行なう。

【0066】第1候補チェック部72は、予測入力機能によって予測された予測文字列が、かな漢字変換部38による入力された読み文字列に対するかな漢字変換処理で得られた同音語候補の第1候補であるか否かを判別する処理を行なう。

【0067】同音語候補入替部74は、第1候補チェック部72によって、予測入力機能によって予測された予測文字列が同音語候補の第1候補でないと判別された場合に、この予測文字列を同音語候補の第1候補として候補順位を入れ替える処理を行なう。

【0068】読み文字列特定部76は、入力されている読み文字列から予測入力に使用する読み文字列を、予測入力用辞書42に登録された情報を参照して特定する処理を行なう。

【0069】読み文字列区別表示部78は、読み文字列特定部76によって特定された予測入力に使用する読み文字列と、その他の読み文字列とを区別して表示する処理を行なう。

【0070】次に、本実施形態における文書作成装置の10動作について、フローチャートを参照しながら説明する。

【0071】はじめに、予測入力機能を用いた基本的な文書作成の手順について説明する。

【0072】ユーザは、かな漢字変換によって文書を作成するためにキーボードなどの入力装置16から読み文字列を入力する。読み文字列の入力途中に、ユーザから予測入力機能を利用した入力の実行指示があると、日本語処理システムは、ユーザによって現在までに入力されている読み文字列をもとにして、ユーザが入力する可能性の高い文字列を予測して、この予測して得られた予測文字列をユーザに提示する。

【0073】ユーザは、日本語処理システムによって画面に提示された予測文字列の中から、入力を希望している文字列を、入力装置16のマウスなどでクリックして直接選択するか、キーボードなどのカーソルキーなどで選択したい文字列にカーソルを移動させリターンキーを押すなどして予測された文字列を読み文字列として選択する。

【0074】本実施形態では、予測入力機能によって予測文字列をユーザに提示する際に、ユーザが入力した読み文字列の情報の他に、現在ユーザがそれまでに入力済みの文書データ、ユーザが以前に入力した文字列の履歴データ等を利用して、ユーザに必要最小限の予測文字列の候補を提示して、予測文字列の候補群からの選択を容易にできるようにしたものである。

【0075】次に、現在作成中の文書の分野を判別して、この文書分野に応じて予測文字列の候補を絞り込む予測入力機能の処理について、図3に示すフローチャートを参照しながら説明する。

【0076】また、図4には、現在作成中の文書の分野を判別して、予測文字列の候補を絞り込み、最適な文字列だけをユーザに提示する具体例を示している。図4(a)はコンピュータ関係の文書1の例、図4(b)は食べ物関係の文書2の例、図4(c)はスポーツ関係の文書3の例を示し、それぞれ同じ読み文字列「そふと」で入力予測を行なった場合に提示される予測文字列を示している。

【0077】まず、ユーザによって入力装置16の予測入力機能の実行を指示する特定のキーが押されるなどすると、入力部32から制御部30に通知されて、予測入

力機能プログラムに従う、日本語処理システムの予測入力機能が起動される（ステップA1）。

【0078】ここでは、図4に示すように、入力部32から「そふと」という読み文字列が入力された時点で、予測入力機能の実行が指示されたものとする。

【0079】予測入力機能が起動されると、制御部30の文書分野特定部52は、現在までに作成されている、データ記憶領域12b（データ記憶部36）によって記憶されている文書のデータをチェックして、文書分野を判別する（ステップA2）。文書分野の特定は、図5に示すような、各単語と各分野の関係を0～2までの数値データ（点数）で関係づけた文書分野特定用辞書44を参照し、以下に示す計算式を使用することで行なう。

【0080】文書分野特定用辞書44には、文書分野毎に文書中で良く使用される単語を示す情報が登録されるもので、例えば図5に示すように、縦方向に配列された分野を特定するための各単語（「最近」「パーソナル」）*

$$\text{文書分野の点数} = \sum (\text{単語の点数} \times \text{単語の重み}) \quad \dots (1)$$

単語の重み

$$= (3 - (\text{全体の行数} - \text{単語のある行})) / (\text{全体の行数} / 3) \quad \dots (2)$$

ただし、分数の分母が1以下の時は1とする。また、1行は句点または改行マークまでの範囲とする。

【0084】文書分野特定部52は、文書内にある全ての単語の各分野での点数の合計を、計算式（1）（2）を用いて求め、点数の高い分野が現在作成中の文書の分野であると判別する。

【0085】計算式（1）（2）では、予測する文字列の近傍に位置する単語が、現在作成中の文書分野に関係の深く、使われる可能性が高い単語である場合が多いという理由から、入力予測の対象とする文字列との相対的な位置関係に応じて重み付けをしている。すなわち、予測する文字列の近傍に位置する単語の点数を位置の遠い単語の点数よりも高くなるように重みづけしている。本実施形態では、単語の位置に応じて単語の点数に1から3の3段階の重みを掛けるようにしている。

【0086】図6には、文書分野特定用辞書44と計算式（1）（2）を用いて、図4（a）の文書1に含まれる各単語から求めた各分野についての点数を示している。この場合、文書1の3行目に入力されている文字列「そふと」についての入力予測を行なっているので、1行目にある単語には重み「1」、2行目の単語には

「2」、3行目の単語には「3」の重みを付けている。

【0087】図6に示すように、文書1については、「計算機」の分野の点数が18点で最も高いので、文書の分野が計算機であると特定できる。

【0088】ここで、前述した計算式（1）（2）を使用する分野の判別が、単純に単語の持つ点数を加算して、点数が最も高い分野を作成中の文書の分野と特定する方法や、単語の持つ点数の平均点から文書分野を特定する方法より優れている例を、図7及び図8に示す。こ

*「コンピュータ」…）の、横方向に配列された分野（「食べ物」「衣服」「計算機」…）における使用される頻度を示すデータが登録されている。本実施形態では、頻度を示すデータとして、良く使われる場合には「2」、使われない場合には「0」、その中間の頻度で使用される場合には「1」の点数が登録されている。

【0081】例えば、単語「コンピュータ」は、「食べ物」の分野で使用されることがないので「0」の点数が登録され、「計算機」の分野では良く使われる所以

10 「2」の点数が登録されている。

【0082】作成中の文書の分野を特定するために使用する計算式は、例えば以下の計算式（1）（2）を用いる。計算式において単語の点数は、図5に示す文書分野特定用辞書44の各単語に対して登録された単語の点数を使用する。

【0083】

$$\text{文書分野の点数} = \sum (\text{単語の点数} \times \text{単語の重み}) \quad \dots (1)$$

単語の重み

$$= (3 - (\text{全体の行数} - \text{単語のある行})) / (\text{全体の行数} / 3) \quad \dots (2)$$

ここで、対象とする文書は、文書の途中で文書分野（文書の話題）が変わるものとする。

【0084】図7は、単語の点数に重みをつけず単純に点数を足した例を示している。

【0085】文書分野が途中で変化する文書の文書分野を特定する場合、単語の点数に重みをつけずに単純に文書中の各単語の点数を加算すると、予測入力の対象としている文字列位置の近傍にある同分野で用いられることが多い単語と、文書分野が変わる以前に入力されている異なる文書分野で用いられる単語とが同じ影響を及ぼすことになる。

【0086】従って、文書分野が変わる以前に入力されている文書分野の単語数が多ければ、異なる文書分野の単語の影響が強くなり、予測入力の対象としている文字列に対して、誤った文書分野を判別しやすくなる。

【0087】図7に示す例では、予測入力の対象としている文字列の近傍（図7に示す下方）に位置する単語に、「スポーツ」の分野の単語が多いにもかかわらず、他に「食べ物」の分野の単語が多く含まれているため

40 に、結局、「スポーツ」の分野の得点が23点であり、「食べ物」の分野の得点が25点で最も高いので、文書分野が「食べ物」と判別される。この場合、「食べ物」関係の文書で使用される例えば「ソフトクリーム」「ソフトドリンク」等の文字列が、入力された読み文字列「そふと」に対する予測文字列として提示されることになる。

【0088】本実施形態では、前述のような誤った文書分野の判別がされないように、単語の位置によって単語の点数に掛ける重みを変え、予測対象とする読み文字列の近傍にある単語の影響を強くすることで、途中で文書

の分野が変わった場合でも分野の特定を誤らないようにしている。

【0094】図8は、単語の点数に重みをつけて点数を計算した例を示している。

【0095】図8に示す例では、予測入力の対象としている文字列の近傍（図7に示す下方）に位置する単語に、重い重みづけをしている。従って、「食べ物」の分野の単語が多く含まれているとしても、「スポーツ」の分野の単語の点数が大きくなるために、結局、「スポーツ」の分野の得点が57点で最も高く、「食べ物」の分野の得点が48点となるので、文書分野が「スポーツ」と判別される。この場合、「スポーツ」関係の文書で使用される例えは「ソフトボール」等の文字列が、入力された読み文字列「そふと」に対する予測文字列として提示されることになる。

【0096】こうして、文書分野特定部52によって、単語の位置に応じて単語の持つ点数に重みづけを行なう文書分野特定方法で文書の分野が判別されると、予測処理部50は、ユーザが現在入力中の読み文字列「そふと」をもとにして、予測入力用辞書42を検索して、入力される可能性が高いと予測できる文字列の候補群を取得する。

【0097】さらに、第1絞り込み部54は、予測処理部50によって予測された文字列の候補群から、文書分野特定部52によって特定された分野の予測文字列のみに絞り込みを行なう（ステップA3）。

【0098】図9には、予測入力用辞書42に登録されている、読み文字列「ソフト」をもとに予測される文字列の候補群についての分野毎使用可能状況データ構造を示している。なお、予測入力用辞書42に登録される文字列には、読み文字列をもとに対応する予測文字列が検出できるように、読みの情報が対応付けられて登録されている。

【0099】図9に示すように、分野毎使用可能状況データには、文書分野毎に各文字列（単語）の使用有無を示すデータ、すなわち良く使われる単語には「1」、使われない単語には「0」が設定されている。第1絞り込み部54は、文書分野特定部52によって特定された分野について、「1」が設定されている文字列（単語）のみを予測文字列として取得する。

【0100】文字列候補表示部56は、第1絞り込み部54によって絞り込まれた予測文字列の候補群を、出力部34を介して表示装置14の画面に表示させて、ユーザに提示する。これにより、ユーザは、予測文字列の候補群から所望する文字列を選択することができる。

【0101】例えば、図4（a）に示すコンピュータ関係の文書1では、「コンピュータ」「ハードウェア」などの文字列から文書1がコンピュータ関係（「計算機」の文書分野）の文書であると判別され、予測文字列の候補群（図9参照）から、「計算機」の文書分野において

「1」が設定された、コンピュータ関係の文書でよく使われる文字列「ソフトウェア」が予測候補としてユーザに提示されている。

【0102】図4（b）に示す食べ物関係の文書2では、「パン」「牛乳」などの文字列から文書2が食べ物関係（「食べ物」の文書分野）の文書であると判別され、予測文字列の候補群（図9参照）から、「食べ物」の文書分野において「1」が設定された、食べ物関係の文書でよく使われる文字列「ソフトクリーム」「ソフトドリンク」が予測候補としてユーザに提示されている。

【0103】図4（c）に示すスポーツ関係の文書3の場合も同様に、「スポーツ」などの文字列から文書3がスポーツ関係（「スポーツ」の文書分野）の文書であると判別され、予測文字列の候補群（図9参照）から、「スポーツ」の文書分野において「1」が設定された、スポーツ関係の文書でよく使われる「ソフトボール」がユーザに提示されている。

【0104】また、図10に示すような異なる文書分野の内容が含まれている文書の場合でも、前述した計算式20（1）（2）と、文書分野特定用辞書44に登録された分野毎使用可能状況データとを利用して、予測の対象とする文字列「そふと」の近傍にある単語の点数に重みをつけて計算を行なうことで、文書分野を正確に判別して、予測文字列の絞り込みを正しく行なうことができる。

【0105】このようにして、予測入力機能を使用して文字列の入力を行なう際に、現在作成中の文書（文書分野）に最適な、必要最小限の予測文字列しか提示しないので、予測文字列の候補群から所望する文字列を選択するための操作を従来に比べて容易にするという効果が得られる。

【0106】なお、前述した説明では、従来の単純に単語の持つ点数を加算して点数が高い分野を作成中の分野と特定する方法や、単語の持つ点数の平均点から文書分野を特定する方法より文書分野特定精度の高い方法として、計算式（1）（2）を使った、入力予測の対象とする文字列の近傍の単語の点数を重みづけをして文書分野の特定の精度を上げる方法を用いたが、文書分野の特定の精度を向上する方法として、単純に文書中の単語の点数の総和による分野判定方法と近傍の単語だけを使った分野判定方法を2つ組み合わせて分野特定を使う方法や、更に文書の表題を文書分野特定の一条件として用いる方法なども考えられる。文書の表題を用いる方法では、例えば文書分野特定用辞書44に文書分野と対応づけて文字列を登録しておき、この文書分野特定用辞書44を参照して、表題に含まれる文字列がどの分野に含まれるかを判別する。

【0107】また、前述した説明では、文書分野特定部52によって文書分野を特定することで予測文字列の候補を絞り込んだが、更にユーザが以前に確定入力した文

字列の履歴情報を予測入力用ユーザ辞書46に登録しておき、この情報を使うことでより一層予測文字列候補を絞り込むことができる。以下、予測入力用ユーザ辞書46に履歴情報を登録する処理について説明する。

【0108】図11は、確定した文字列を予測入力用ユーザ辞書44に登録する処理の流れを示すフローチャートである。

【0109】日本語処理システムは、ユーザの操作によって入力された文字列を文書分野特定用辞書44に登録していく。

【0110】日本語処理システムの機能を用いて文字列が入力されると、確定文字列解析部64は、文字列解析用辞書48に登録された情報を参照しながら解析処理を行ない、入力された文字列を予測入力用辞書42に登録されるルートである自立語に分解する(ステップB1)。

【0111】例えば、入力部32から読み文字列が入力され、かな漢字変換部38によってかな漢字変換辞書40に登録された情報を参照しながらかな漢字変換が行われ、複数の変換候補が得られたものとする。ここで、入力する文字列を選択する指示が入力部32から入力されて確定されると、この確定された文字列が確定文字列解析部64による解析の対象となる。

【0112】確定文字列解析部64による解析が行われた後、予測入力用ユーザ辞書処理部58は、自立語に分解された各文字列について、既に予測入力用ユーザ辞書46に登録されているか否かを判別する(ステップB2)。

【0113】ここで、確定された文字列が予測入力用ユーザ辞書46に登録されていない場合には、予測入力用ユーザ辞書処理部58は、文字列記憶部58aによって、その文字列を予測入力用ユーザ辞書46に登録する(ステップB3)。

【0114】本実施形態では、予測入力用ユーザ辞書46に、図12に示すように、確定された文字列の読み文字列と、入力された確定文字列と、使用頻度(使用回数)を示す頻度情報とが対応づけられて登録されるものとする。

【0115】図12に示す例では、読み文字列「そふとうえあ」に対して確定文字列「ソフトウェア」、読み文字列「そふとくりーむ」に対して確定文字列「ソフトクリーム」、読み文字列「そふとぼーる」に対して確定文字列「ソフトボール」が登録されている例である。

【0116】一方、確定された文字列が既に予測入力用ユーザ辞書46に登録されている場合には、予測入力用ユーザ辞書処理部58は、頻度情報記憶部58bによって、その文字列に対応する頻度情報に1加算し、頻度情報を更新する(ステップB4)。

【0117】次に、前述のようにして文字列が登録された予測入力用ユーザ辞書46を用いた予測入力機能を利用した文字列の入力について説明する。

【0118】図13は、予測入力用ユーザ辞書46を使用して、予測文字列の候補を絞り込む処理の流れを示すフローチャートである。

【0119】まず、ユーザによる予測入力機能の実行の指示が入力部32から入力されると(ステップC1)、制御部30の文書分野特定部52は、現在までに作成されている、データ記憶領域12b(データ記憶部36)によって記憶されている文書のデータをチェックして、文書分野を判別する(ステップC2)。なお、文書分野特定部52による文書分野の判別の処理は、図3のステップA2において説明した処理と同様にして行われるものとして説明を省略する。

【0120】また、予測処理部50は、現在入力中の読み文字列をもとにして、分野毎使用可能状況データが登録された予測入力用辞書42を検索して、入力される可能性が高いと予測できる文字列の候補群を取得する。さらに、第1絞り込み部54は、予測処理部50によって予測された文字列の候補群から、文書分野特定部52によって特定された分野の予測文字列のみに絞り込みを行なう(ステップC3)。なお、ステップC3の処理は、図3のステップA3において説明した処理と同様にして行われるものとして詳細な説明を省略する。

【0121】この後、予測文字列が予測入力用ユーザ辞書46に登録されている文字列かどうかを判別し(ステップC4)、予測入力用ユーザ辞書46に登録されている場合には、第2絞り込み部60は、その文字列だけを予測文字列の候補として絞り込む。文字列候補並べ替え部62は、第2絞り込み部60によって絞り込まれた予測文字列が複数ある場合には、複数の予測文字列の順番を、予測入力用ユーザ辞書46に登録された文字列とその頻度情報に基づいて並べ替える。すなわち、頻度の値が大きい予測文字列ほど、複数の予測文字列からの選択が容易となるように上位に順番を変更する。文字列候補表示部56は、第2絞り込み部60によって絞り込まれ、文字列候補並べ替え部62によって順番が並べ替えられた予測文字列を表示してユーザに提示する(ステップC5)。

【0122】一方、予測入力用ユーザ辞書に登録されていない場合には、文字列候補表示部56は、ステップC3において取得された予測文字列を、そのまま表示してユーザに提示する(ステップC6)。

【0123】前述した例では、予測入力用ユーザ辞書46に登録された文字列だけをユーザに提示するようにしたが、ユーザに提示する予測文字列の数を同じにして、予測入力用ユーザ辞書46に各文字列に対応して登録されている頻度情報を使って頻度が高いものから順番に必要な数の文字列を取得するようにもできる。

【0124】この場合、文字列候補並べ替え部62は、前述と同様にして、頻度情報を使って頻度が高いものから順番に予測文字列を並べ替えて、文字列候補表示部5

6によって表示させる。

【0125】図14には、図12に示す予測入力用ユーザ辞書46に登録された文字列をもとにして、第2絞り込み部60による予測文字列の候補の絞り込みが行われた例を示している。

【0126】読み文字列「そふと」の入力中に予測入力機能による予測入力の指示があると、文書分野特定部52によって作成中の文書中の単語をもとにして文書分野を「食べ物」関係であると判別され、食べ物関係の文書によく使われる「ソフトクリーム」「ソフトドリンク」が予測文字列の候補として取得されたものとする。

【0127】ここで、図12に示すような情報が登録された予測入力用ユーザ辞書46を参照とすると、文字列「ソフトクリーム」が登録されていることが判別される。

【0128】第2絞り込み部60は、複数の予測文字列から予測入力用ユーザ辞書46に登録された文字列と同じ予測文字列「ソフトクリーム」のみを取得する。文字列候補表示部56は、第2絞り込み部60によって取得された予測文字列を、図14に示すようにして表示させてユーザに提示する。

【0129】このようにして、ユーザが以前入力した文字列を記憶している予測入力用ユーザ辞書46を利用することで、ユーザの使用環境に応じて予測文字列を更に絞り込むことができるようになる。従って、予測入力機能によって提示された予測文字列の候補群から所望する文字列を選択するための操作をより容易にすることができる。

【0130】なお、前述した説明では、予測入力用ユーザ辞書46に登録される文字列として、かな漢字変換によって確定入力された文字列としているが、予測入力機能によって提示された予測文字列から入力すべき文字列として選択された文字列を登録することも可能である。

【0131】次に、現在作成中の文書分野であまり使わない文字列をユーザが入力したい場合にも、ユーザの要求に柔軟に応えることができるようするための機能について説明する。

【0132】図15は、予測入力機能によって提示された予測文字列以外の文字列を入力すべき文字列として確定した場合に、予測入力用辞書42に登録された分野毎使用可能状況データを更新する処理の流れを示すフローチャートである。図16に示す、予測入力機能によって提示された予測文字列以外の文字列が確定入力された場合の具体例を参照しながら説明する。

【0133】ここでは、図16に示すように、コンピュータ関係の文書を入力中にユーザが「ソフトクリーム」というソフトウェアの製品名を入力しようとして、読み文字列「そふと」を入力した時点で、予測入力機能による入力予測の実行が指示されたものとする。

【0134】この場合、図11、図13を用いて説明し

たような予測入力機能によって、作成中の文書がコンピュータ関係の文書であるので読み文字列「そふと」をもとにして取得された予測文字列のうち、図16(a)に示すように、コンピュータ関係の文書分野でよく使われる文字列「ソフトウェア」だけが予測候補としてユーザに提示されたものとする。

【0135】この場合、ユーザが入力したいのは「ソフトクリーム」であるため、提示された予測文字列「ソフトウェア」が入力に使用されずに、図16(b)に示すように、続けて読み文字列が入力され、かな漢字変換が実行されたものとする、すなわち、図16(c)に示すように、「そふとくりーむ」を含む読み文字列についての変換文字列として「ソフトクリームというソフトウェアである。」が確定入力されたものとする。

【0136】文字列が確定入力されると、確定文字列解析部64は、文字列解析用辞書48に登録された情報を参照しながら解析処理を行ない、入力された文字列を自立語に分解する(ステップD1)。

【0137】辞書判別部66は、確定文字列解析部64による解析で得られた自立語(確定入力された文字列)が予測文字列の候補群にあるかどうかを判別する(ステップD2)。すなわち、辞書判別部66は、確定入力された文字列が、文書の分野情報を使って予測文字列の候補を絞る前の予測文字列の候補群(図9に示す読み文字列「そふと」に対応する予測文字列)に含まれる文字列であるか否かを判別する。

【0138】ここで、確定入力された文字列が、予測文字列に含まれていた場合には、辞書判別部66は、その文字列が現在作成中の文書の分野であまり使われない文字列であるか否かをチェックする(ステップD3)。すなわち、辞書判別部66は、予測入力用辞書42に登録された分野毎使用可能状況データを参照して、現在作成中の文書の文書分野における、対象とする単語について設定されたデータ(点数)が、使用されることを示す「1」であるか、使用されない「0」であるかによって判別する。

【0139】ここで、対象とする単語に対して、使用されないことを示す「0」が設定されていた場合には、予測入力用辞書処理部68の内容変更部68bは、確定入力された文字列の持つ現在の分野に対応するデータを「0」から「1」に変更し、対象とする単語を現在作成中の文書の分野でよく使われる単語となるようにする(ステップD4)。

【0140】例えば、図16に示すように、コンピュータ分野ではあまり使われない文字列「ソフトクリーム」がコンピュータ関係の文書の作成中に確定入力された場合の、分野毎使用可能状況データの変化の例を図17に示している。前述した例では、図17に示すように、図17(a)に示す「計算機」の文書分野の単語「ソフトクリーム」に対する点数が「0」であったものを、図1

7 (b) に示すように、「1」に更新されている。

【0141】内容変更部68bによるデータの更新によって、次回ユーザがコンピュータ分野の文書を作成中に読み文字列「そふと」で予測を指示すると、予測文字列「ソフトウェア」の他にユーザが前回確定入力した「ソフトクリーム」も予測文字列の候補として追加されて提示される。

【0142】一方、ステップD2において、確定入力された文字列が予測文字列の候補群に含まれないと判別された場合、予測入力用辞書処理部68の文字列記憶部68aは、予測入力用辞書42に登録された分野毎使用可能状況データに、確定入力された文字列を新しく追加する(ステップD5)。これにより、次回ユーザが同じ読み文字列を入力して、予測入力機能を利用した入力を行なう際に、前回ユーザが確定入力した文字列も予測文字列の候補に追加されて提示されるようになる。

【0143】例えば、読み文字列「そふと」をもとにした入力予測によって予測文字列として取得されなかつた「ソフトハウス」が確定入力された場合、図18(a)に示すように、「ソフトハウス」の文字列が登録されていない分野毎使用可能状況データに対して、内容変更部68bは、確定入力された「ソフトハウス」を図18(b)に示すように新しく追加する。この場合、「計算機」の文書分野における「ソフトハウス」に対する点数が「1」に設定されている。

【0144】このようにして、日本語処理システムは、ユーザが確定入力した文字列をチェックしながら、予測入力用辞書42に登録された分野毎使用可能状況データの内容を更新する機能を持つことで、予測入力機能を用いて文字列を入力する際に提示される予測文字列を、文書を作成するほど、ユーザ毎、作成する文書の文書分野毎に最適になるように変えていくことができるようになる。

【0145】次に、予測入力機能によって得られた予測文字列をかな漢字変換に利用して、かな漢字変換結果の変換率向上を図る場合について説明する。

【0146】図19は、入力予測された文字列をそのままかな漢字変換の読み文字列とするかな漢字変換処理の流れを示すフローチャートである。

【0147】まず、かな漢字変換部38によるかな漢字変換によって文字列を入力するために、入力部32から読み文字列が入力され、この読み文字列をもとに予測入力機能を利用して文字列が入力されたものとする。この場合、漢字を含む予測文字列が選択されて入力された場合も、そのまま漢字も含めて読み文字列として扱われる。

【0148】ここで、かな漢字変換の変換実行が指示されると、かな漢字変換部38は、読み文字列中に、予測文字列が含まれているかどうかを、予測入力用辞書42に登録された情報を参照することによって判別する(ス

テップE1)。

【0149】読み文字列中にユーザが選択した予測文字列が含まれている場合、かな漢字変換部38は、読み文字列に含まれる予測文字列を利用し、かな漢字変換辞書40に登録されている共起情報を参照することによって、同音語候補を絞ったかな漢字変換を行なう(ステップE2)。

【0150】一方、読み文字列中に予測文字列が含まれない場合、かな漢字変換部38は、通常のかな漢字変換を行なう(ステップE3)。

【0151】図20には、予測文字列をそのままかな漢字変換の読み文字列として利用したかな漢字変換の一例を示している。

【0152】まず、図20(a)に示すように、読み文字列「よこはま」が入力された時点で予測入力機能による入力の実行が指示され、前述したような方法によって予測文字列の候補がユーザに提示され、この候補から図20(b)に示すように「横浜ベイブリッジ」の文字列が選択されたものとする。

【0153】予測文字列「横浜ベイブリッジ」が選択入力された後、図20(c)に示すように、そのまま読み文字列「をかける」が入力された後に、かな漢字変換の実行が指示されたものとする。

【0154】ここで、かな漢字変換部38は、かな漢字変換において、予測文字列「横浜ベイブリッジ」を含んだ読み文字列を「横浜ベイブリッジを／かける」というように文節切りを行なう(／が文節の切れ目)。

【0155】かな漢字変換部38は、かな漢字変換辞書40に登録された単語情報をもとにして得られる「かける」に対する複数の同音語候補(例えば、架ける、欠ける、掛ける、駆ける等)の中から、共起情報を参照して、「ブリッジ」(橋、建築物等の分野)と共に共起関係にある「架ける」に候補を絞って、かな漢字変換の結果として「架ける」を出力する。

【0156】このようにして、予測文字列をそのまま読み文字列としてかな漢字変換を行なう場合は、ユーザが選択した予測文字列が読み文字列の中のどの部分であるかわかり、また予測文字列部分を確定した文字列(曖昧性のない確定文字列情報)として扱うことができるため、この情報を用いて同音語候補の絞り込みを行なうことで変換結果の向上を図ることができる。

【0157】前述では、かな漢字変換結果の変換率向上を図るために、漢字を含む予測文字列をそのまま読み文字列として使用するかな漢字変換部38を用いるものとして説明したが、かな漢字変換部38がひらがなによる読み文字列のみを対象とする場合の予測文字列を優先したかな漢字変換処理について説明する。

【0158】図21は、ユーザが選択した予測文字列を優先したかな漢字変換処理の流れを示すフローチャートである。ここでは、選択された予測文字列を一度ひらが

なの読み文字列に戻してかな漢字変換を行ない、以下の処理を行なうことで、ユーザが望む予測文字列をかな漢字変換の結果として取得できるようになる。

【0169】まず、かな漢字変換部38によるかな漢字変換によって文字列を入力するために、入力部32から読み文字列が入力され、この読み文字列をもとに予測入力機能を利用して予測文字列が選択されたものとする。ここで、選択された予測文字列については、予測入力用辞書42に登録されている該当する予測文字列に対応する読みの情報をもとにして、ひらがなの読みに変換して、かな漢字変換部38によるかな漢字変換の対象となる読み文字列とする。この予測文字列から得られた読み文字列に対しては、続けて読み文字列を入力したり、変更したりすることができる。

【0170】ここで、かな漢字変換の変換実行が指示されると、かな漢字変換部38は、読み文字列中に、ユーザによって選択された予測文字列の読みが含まれているかどうかを、予測入力用辞書42に登録された情報を参照することによって判別する(ステップF1)。

【0171】読み文字列中にユーザが選択した予測文字列の読みが含まれている場合は、かな漢字変換部38は、ユーザによって選択された予測文字列の読み文字列に対する文節区切りと同じ位置を優先した文節切りを行ない、かな漢字変換辞書40に登録されている情報を参照することによって、かな漢字変換を実行する(ステップF2)。

【0172】さらに、かな漢字変換部38によるかな漢字変換の実行の後、同音語候補チェック部70は、かな漢字変換によって得られた同音語候補中に、予測文字列と同じ読み文字列を持つ同音語があるかどうか、予測入力用辞書42を参照しながら判別する。

【0173】ここで、予測文字列の読みと同じ同音語候補がある場合は、第1候補チェック部72は、該当する予測文字列がその同音語候補中の第1候補であるかを判別する(ステップF4)。

【0174】この結果、予測文字列が同音語候補中の第1候補でない場合は、同音語候補入替え部74は、ユーザが選択した予測文字列を、かな漢字変換による変換結果である同音語候補中の第1候補になるように変更する(ステップF5)。

【0175】一方、読み文字列中に予測文字列が含まれない場合、かな漢字変換部38は、通常のかな漢字変換を行なう(ステップF6)。

【0176】図22には、予測文字列を優先したかな漢字変換の一例を示している。

【0177】まず、図22(a)に示すように、読み文字列「じどう」が入力された時点で予測入力機能による入力の実行が指示され、前述したような方法によって予測文字列の候補、例えば「児童送迎バス」がユーザに提示され、この予測文字列が選択されたものとする。

【0168】この場合、予測入力用辞書42にほ予測文字列「児童送迎バス」に対応して読み「じどうそうげいばす」が登録されており、図22(b)に示すように、この読み「じどうそうげいばす」が読み文字列として扱われる。

【0169】図22(b)に示す読み文字列に対してかな漢字変換を行なうために、通常では、かな漢字変換部38は、かな漢字変換辞書40に登録された情報を使って「じ／どうそ／うげ／いばす」「じどうそ／げいばす」「じどう／そうげい／ばす」など様々な位置で読み文字列の分割を行なう。

【0170】本実施形態では、予測入力機能を用いて予測文字列が選択されている場合、かな漢字変換部38は、ユーザが選択した予測文字列「児童送迎バス」のそれぞれの文字に対応する読み文字列「じ」「どう」「そ」「う」「げい」「ば」「す」を使ってかな漢字変換の辞書にある最長の単語を求め、その単位で読み文字列を区切るようにする。

【0171】具体的には、最初に「じ」という読みでかな漢字変換辞書40を調べると「次」「時」などいくつかの単語がある。次に「じどう」で調べると、「自動」「児童」という単語があるので、さらに「じどうそう」で調べるとこの読みに対応する単語がないということで、最初の切れ目は「じどう」の後にくるという判別がされる。

【0172】この処理を繰り返すことで「じどうそうげいばす」の読み文字列の切れ目は「じどう／そうげい／ばす」と判別される。かな漢字変換部38は、この読み文字列の切れ目の位置を示す情報を用い、切れ目によって区切られる読み文字列毎にかな漢字変換処理を行なう。

【0173】この例では、かな漢字変換によって得られた同音語候補として、例えば、読み文字列「じどう」に対しては第1候補「自動」、第2候補「児童」が得られ、読み文字列「そうげい」に対しては第1候補「送迎」、第2候補「創芸」が得られる。

【0174】この結果、共起情報等に基づいて、変換結果とする同音語候補としては第1候補「自動送迎バス」、第2候補「児童送迎バス」の2つが得られる。通常の変換結果としては、第1候補「自動送迎バス」が 출력される。

【0175】従って、第1候補チェック部72によって、予測文字列「児童送迎バス」が2つの同音語候補中の第1候補でないことが判別されるため、同音語候補入替え部74は、ユーザが選択した予測文字列「児童送迎バス」が同音語候補中の第1候補になるように変更する。この場合、予測文字列中の「児童」が「自動」よりも低い優先順位であるので、「児童」の優先順位を上げることになる。この結果、図22(c)に示すように、「児童送迎バス」がかな漢字変換部38によるかな漢字

変換の結果として出力される。

【0176】このようにして、かな漢字変換部38が、通常のひらがなによる読み文字列のみを対象とするかな漢字変換処理を行なう場合に、かな漢字変換辞書40に登録された情報をもとにしたかな漢字変換では「自動送迎バス」という変換結果が第1候補となる場合も、予測文字列をひらがなに戻してかな漢字変換を行ない、ユーザによって選択された予測文字列を優先させるかな漢字変換を行なうことによって、「児童送迎バス」を変換結果として出力できるようになる。

【0177】次に、複数文節の読み文字列が入力された際に予測入力機能を利用する場合について説明する。

【0178】図23は、読み文字列が複数文節入力された際の予測入力に使用する文字列範囲を他の文字列と区別して表示するための処理の流れを示すフローチャートである。

【0179】複数文節の読み文字列を入力中に、ユーザが予測入力の指示を行ないたい時などに、予測入力に使用する読み文字列の範囲が何も示されていない場合には、予測入力の実行を指示した際に、どの読み文字列に対する予測入力が行なわれるのかユーザにとって分かりづらい。そこで、予測入力に使用される読み文字列の範囲を、読み文字列の入力にしたがって他の読み文字列と区別して表示することで、予測入力に使われる読み文字列がユーザにとってわかり易くする。

【0180】まず、入力部32からかな漢字変換により文書を作成するために読み文字列が入力されると、順次、かな漢字変換部38によるかな漢字変換の対象とする通常の読み文字列として表示される。一方、読み文字列特定部76は、予測入力用辞書42を参照しながら、入力された文字を含む読み文字列から予測入力に使用する読み文字列の範囲を特定する(ステップG1)。

【0181】すなわち、入力された文字を含む読み文字列から、読み文字列特定部76は、予測入力用辞書42に登録されている少なくとも1つの予測文字列が特定できるか否かを判別し、該当する予測文字列が存在する場合に、入力された読み文字列を入力予測の範囲(1つ文節の範囲に該当する)として特定する。

【0182】なお、予測入力用辞書42に登録された情報を利用せずに、単に文節の区切り位置を判別し、直前に入力された文字を含む文節に該当する文字列を、予測入力の対象範囲として特定するようにしても良い。

【0183】読み文字列特定部76によって予測入力に使用される読み文字列が特定できた場合(ステップG2)、読み文字列区別表示部78は、この予測入力に使用する読み文字列の範囲を、他の読み文字列と区別できる表示属性を付加して表示させる(ステップG3)。

【0184】図24には、複数文節の読み文字列が順次入力される場合の一例を示している。

【0185】図24(a)に示すように、読み文字列

「たいいく」が入力された時点で予測入力が可能な範囲であると判別され、この文字列の範囲に下線が付加されて、予測入力の範囲が明示される。ここで、予測入力の実行が指示されると、下線が付加された読み文字列「たいいく」をもとに、前述した方法による予測入力機能による処理により予測文字列が提示される。

【0186】また、図24(b)に示すように、読み文字列「たいいく」に続けてさらに「のじかん」と文字列が入力された場合、文字列「じかん」が入力された時点で予測入力が可能な範囲であると判別され、先に下線が付加されている文字列「たいいく」に替えて下線が付加される。ここで、予測入力の実行が指示されると、下線が付加された読み文字列「じかん」をもとにした予測文字列が提示される。

【0187】さらに、図24(c)に示すように、読み文字列「じかん」に続けてさらに「はそふと」と文字列が入力された場合、文字列「そふと」が入力された時点で予測入力が可能な範囲であると判別され、先に下線が付加されている文字列「じかん」に替えて下線が付加される。ここで、予測入力の実行が指示されると、下線が付加された読み文字列「そふと」をもとにした予測文字列が提示される。

【0188】このようにして、入力される読み文字列に応じて、予測入力の対象とする文字列が判別されて、該当する文字列の範囲にのみ、順次、他と異なる表示属性が付加されて表示されるので、ユーザは、予測入力機能の実行を指示した際に、予測入力の対象となる読み文字列を容易に把握することができる。

【0189】なお、図24に示す例では、予測入力の対象となる読み文字列の範囲を区別するための表示形態として文字列に下線を付加する方法をとっているが、他の表示形態として、該当する文字列の各文字のフォントを変えたり、文字のサイズを変えたり、文字の表示色を変えたり、斜体にするなど、ユーザが予測に使う読み文字列を他の読み文字列と区別できる方法を利用することができる。

【0190】なお、上述した実施形態において記載した各手法は、コンピュータに実行させることのできるプログラムとして、例えば磁気ディスク(フロッピーディスク、ハードディスク等)、光ディスク(CD-ROM、DVD等)、半導体メモリなどの記録媒体に書き込んで各種装置に適用したり、通信媒体により伝送して各種装置に適用することも可能である。本装置を実現するコンピュータは、記録媒体に記録されたプログラムを読み込み、このプログラムによって動作が制御されることにより、上述した処理を実行する。

【0191】

【発明の効果】以上詳述したように本発明によれば、現在作成中の文書内の文字列情報と文字列の位置情報を使って文書分野の特定を行なうことで文書分野特定の精度

を高め、この方法で特定した文書分野を使って予測文字列候補を絞り込むことで、ユーザに対してはよく使われる最適な文字列だけを提示できるようになり、表示された予測文字列からユーザが選択する操作を簡単にすることができる。

【図面の簡単な説明】

【図1】本発明の実施形態に係わる文書作成装置のシステム構成を示すブロック図。

【図2】本実施形態における文書作成装置の機能構成を示すブロック図。

【図3】現在作成中の文書の分野を判別して、この文書分野に応じて予測文字列の候補を絞り込む予測入力機能の処理を説明するためのフローチャート。

【図4】現在作成中の文書の分野を判別して、予測文字列の候補を絞り込み、最適な文字列だけをユーザに提示する具体例を示す図。

【図5】文書分野特定用辞書44に登録されたデータの内容を説明するための図。

【図6】文書分野特定用辞書44と計算式(1)(2)を用いて、図4(a)の文書1に含まれる各単語から求めた各分野についての点数を示す図。

【図7】文書分野を特定するための処理において、単語の点数に重みをつけて単純に点数を足した例を示す図。

【図8】文書分野を特定するための処理において、単語の点数に重みをつけて点数を計算した例を示す図。

【図9】予測入力用辞書42に登録されている、読み文字列「ソフト」をもとに予測される文字列の候補群についての分野毎使用可能状況データ構造を示す図。

【図10】異なる文書分野の内容が含まれている文書の一例を示す図。

【図11】本実施形態における確定した文字列を予測入力用ユーザ辞書44に登録する処理の流れを示すフローチャート。

【図12】予測入力用ユーザ辞書46に登録された、確定された文字列の読み文字列と、入力された確定文字列と、使用頻度(使用回数)を示す頻度情報とが対応づけられた一例を示す図。

【図13】本実施形態における予測入力用ユーザ辞書46を使用して、予測文字列の候補を絞り込む処理の流れを示すフローチャート。

【図14】図12に示す予測入力用ユーザ辞書46に登録された文字列をもとに、第2絞り込み部60による予測文字列の候補の絞り込みが行われた例を示す図。

【図15】予測入力機能によって提示された予測文字列以外の文字列を入力すべき文字列として確定した場合に、予測入力用辞書42に登録された分野毎使用可能状況データを更新する処理の流れを示すフローチャート。

【図16】予測入力機能によって提示された予測文字列以外の文字列が確定入力された場合の具体例を示す図。

【図17】文字列「ソフトクリーム」がコンピュータ関

係の文書の作成中に確定入力された場合の分野毎使用可能状況データの変化の例を示す図。

【図18】「ソフトハウス」の文字列が登録されていない分野毎使用可能状況データの更新の例を示す図。

【図19】本実施形態における入力予測された文字列をそのままかな漢字変換の読み文字列とするかな漢字変換処理の流れを示すフローチャート。

【図20】予測文字列をそのままかな漢字変換の読み文字列として利用したかな漢字変換の一例を示す図。

10 【図21】ユーザが選択した予測文字列を優先したかな漢字変換処理の流れを示すフローチャート。

【図22】予測文字列を優先したかな漢字変換の一例を示す図。

【図23】読み文字列が複数文節入力された際の予測入力に使用する文字列範囲を他の文字列と区別して表示するための処理の流れを示すフローチャート。

【図24】複数文節の読み文字列が順次入力される場合の一例を示す図。

20 【図25】従来の予測入力機能を用いた文字入力例を示す図。

【図26】従来の予測文字列から選択した文字列に続けて読み文字を入力し、かな漢字変換を行なう状況を説明するための図。

【符号の説明】

1 0 … C P U

1 2 … メモリ

1 4 … 表示装置

1 6 … 入力装置

1 8 … 出力装置

30 2 0 … 外部記憶装置

3 0 … 制御部

3 2 … 入力部

3 4 … 出力部

3 6 … データ記憶部

3 8 … かな漢字変換部

4 0 … かな漢字変換辞書

4 2 … 予測入力用辞書

4 4 … 文書分野特定用辞書

4 6 … 予測入力用ユーザ辞書

40 4 8 … 文字列解析用辞書

5 0 … 予測処理部

5 2 … 文書分野特定部

5 4 … 第1絞り込み部

5 6 … 文字列候補表示部

5 8 … 予測入力用ユーザ辞書処理部

5 8 a … 文字列記憶部

5 8 b … 頻度情報記憶部

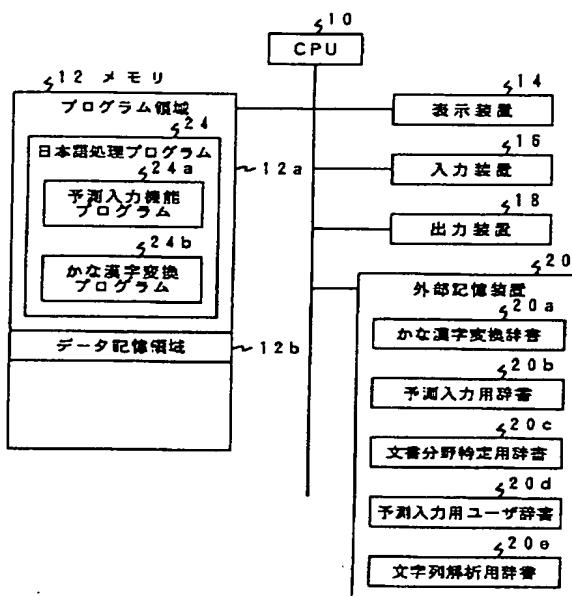
6 0 … 第2絞り込み部

6 2 … 文字列候補並べ替え部

6 4 … 確定文字列解析部

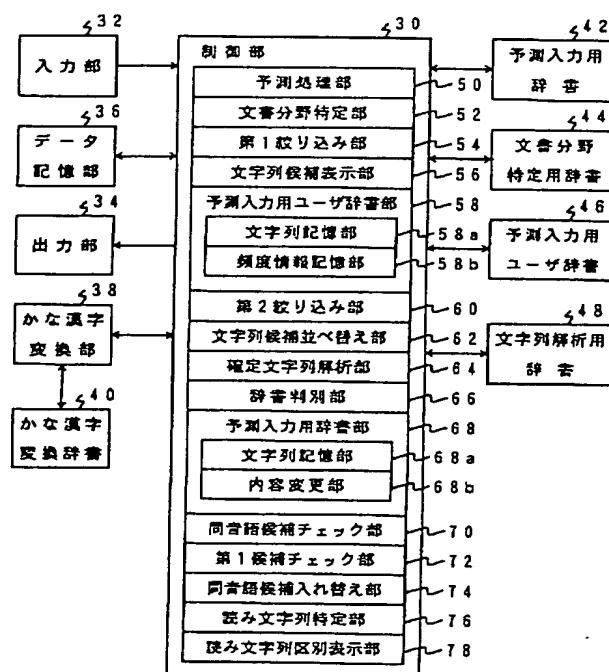
6 6 … 辞書判別部
 6 8 … 予測入力用辞書処理部
 6 8 a … 文字列記憶部
 6 8 b … 内容変更部
 7 0 … 同音語候補チェック部

【図1】

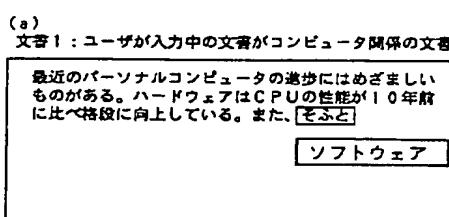


7 2 … 第1候補チェック部
 7 4 … 同音語候補入替え部
 7 6 … 読み文字列特定部
 7 8 … 読み文字列区別表示部

【図2】



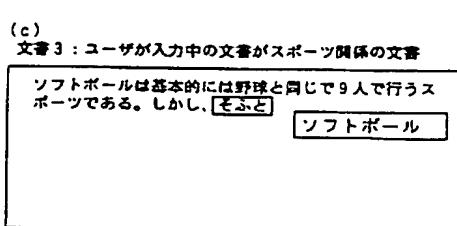
【図4】



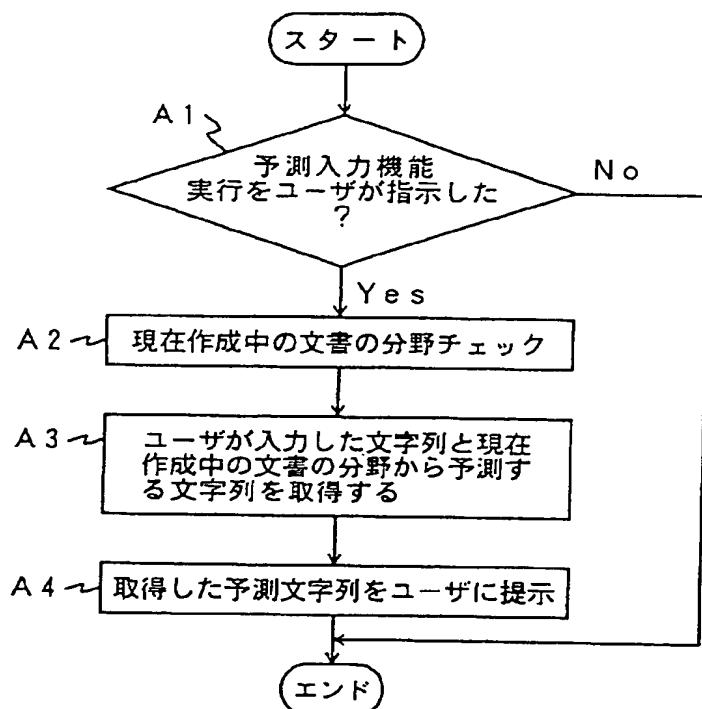
【図6】

	食べ物	衣服	計算機	スポーツ	…	政治
最近	1	1	1	1		1
パーソナル	1	1	2	1		0
コンピュータ	0	0	2	0		0
進歩	1	1	1	1		1
ハードウェア	0	0	4	0		0
CPU	0	0	4	0		0
性能	2	2	2	2		2
格段	2	2	2	2		2
合計	7	7	18	7	…	6

作成中文書の分野を特定するための計算結果例



【図3】



【図5】

	食べ物	衣服	計算機	スポーツ	…	政治
最近	1	1	1	1		1
パーソナル	1	1	2	1		0
コンピュータ	0	0	2	0		0
進歩	1	1	1	1		1
ハードウェア	0	0	2	0		0
CPU	0	0	2	0		0
性能	1	1	1	1		1
格段	1	1	1	1		1
ボール	2	1	0	2		0
料理	2	0	0	0		0
鉢	1	0	0	0		0
材質	1	1	0	1		0
金属	1	1	1	1		0
ガラス	1	1	0	0		0
木	1	1	0	1		0
陶器	1	0	0	0		0
肉	2	0	0	0		0
魚	2	0	0	0		0
野菜	2	0	0	0		0
素材	1	1	0	1		0
野球	0	0	0	2		0
チーム	0	0	0	2		0
スポーツ	1	1	0	2		0
バット	0	0	0	2		0
グローブ	0	1	0	2		0
道具	1	1	1	1		1
…						

文書分野の特定のための辞書構造例

【図7】

単語の点数に重みをつけて単純に点数を足した場合

	食べ物	衣服	計算機	スポーツ	…	政治
ボール	6	3	0	6		0
料理	4	0	0	0		0
鉢	1	0	0	0		0
材質	1	1	0	1		0
金属	1	1	1	1		0
ガラス	1	1	0	0		0
木	1	1	0	1		0
陶器	1	0	0	0		0
肉	2	0	0	0		0
魚	2	0	0	0		0
野菜	2	0	0	0		0
素材	1	1	0	1		0
野球	0	0	0	4		0
チーム	0	0	0	2		0
スポーツ	1	1	0	2		0
バット	0	0	0	2		0
グローブ	0	1	0	2		0
道具	1	1	1	1		1
合計	25	11	2	23		1

【図8】

単語の点数に重みをつけて点数を計算した場合

	食べ物	衣服	計算機	スポーツ	…	政治
ボール	10	5	0	10		0
料理	6	0	0	0		0
鉢	2	0	0	0		0
材質	2	2	0	2		0
金属	2	2	2	2		0
ガラス	2	2	0	0		0
木	2	2	0	2		0
陶器	2	0	0	0		0
肉	4	0	0	0		0
魚	4	0	0	0		0
野菜	4	0	0	0		0
素材	2	2	0	2		0
野球	0	0	0	12		0
チーム	0	0	0	6		0
スポーツ	3	3	0	6		0
バット	0	0	0	6		0
グローブ	0	3	0	6		0
道具	3	3	3	3		3
合計	48	24	5	57		3

[図9]

	食べ物	衣服	計算機	スポーツ	...	政治
ソフト	1	1	1	1		1
ソフトウェア	0	1	1	0		0
ソフトボール	0	0	0	1		0
ソフトクリーム	1	0	0	0		0
ソフトドリンク	1	0	0	0		0
ソフトカラー	0	1	0	0		0
ソフトフォーカス	0	0	0	0		0
...						

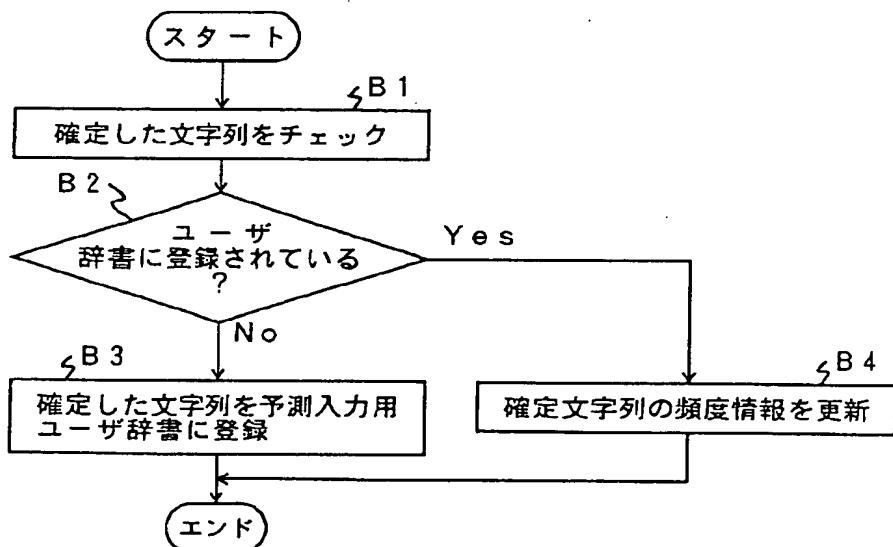
分野毎使用可能状況データ構造例

[図 10]

以下の文では、同じ「ボール」でも違うものを指している。

- ・料理で使うボールは丸形の深い鉢である。
- ・材質は、金属性・ガラス製・木製・陶製など種類が多い。
- ・このボールを使い、料理に応じて肉、魚、野菜など様々な素材を混ぜ合わせる。
- ・野球は1チーム9人で行うスポーツであり、ボール、バット、グローブなどの道具を使って行う。野球に似たスポーツとしては、**【モフド】**

【图 1-1】



[図 1-2]

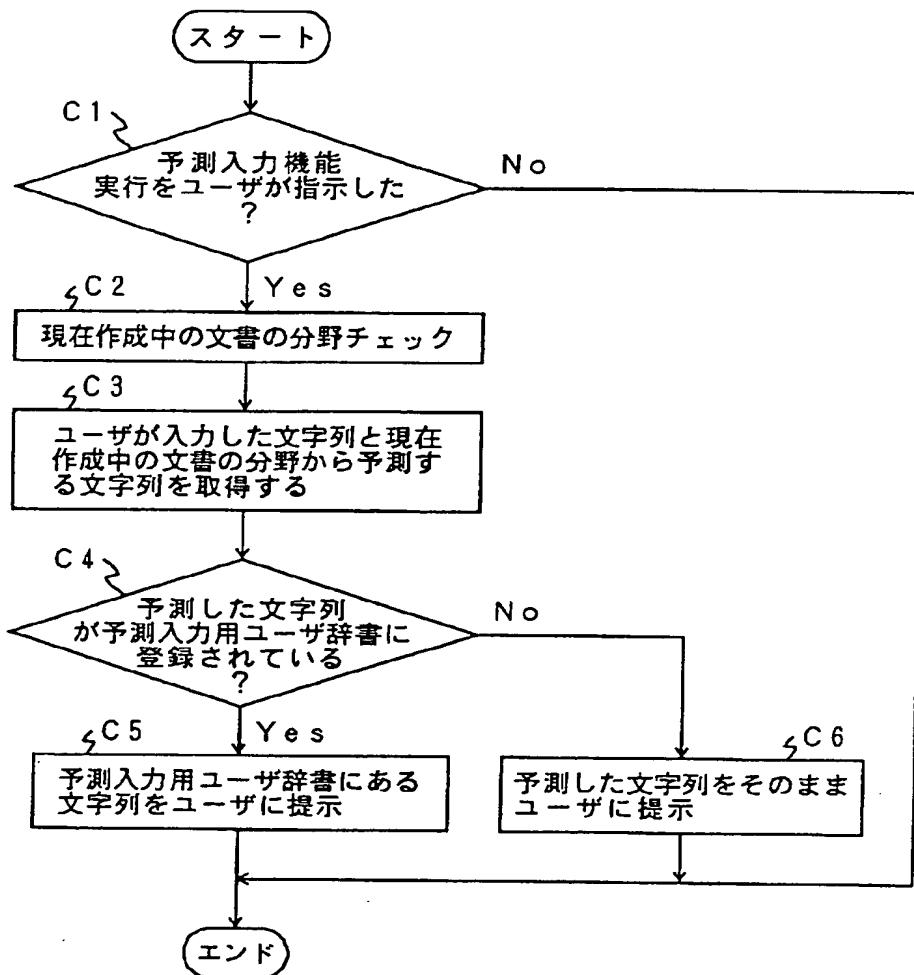
確定文字列の読み	確定文字列	頻度
…	…	
そふとうえあ	ソフトウェア	1
そふとくりーむ	ソフトクリーム	1
そふとぼーる	ソフトボール	1
…	…	

予測入力用ユーザ辞書内容例

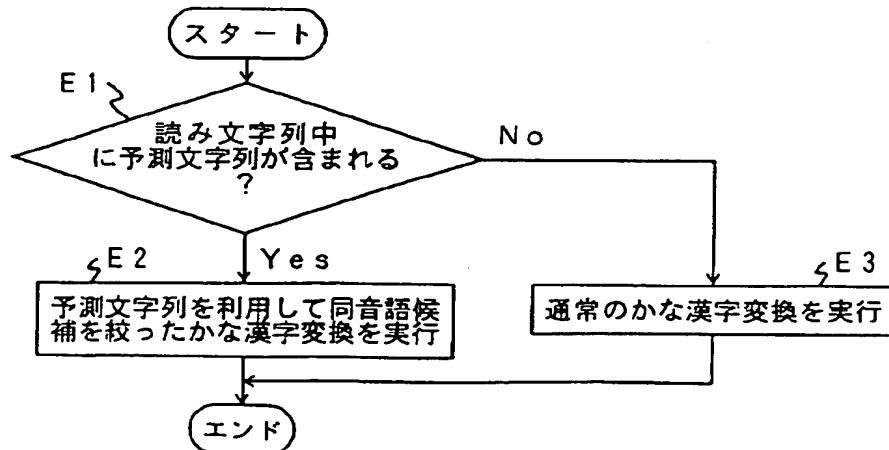
[図14]

2月26日の食事
朝食：パン、牛乳
昼食：うどん
夕食：焼き魚、ご飯、みそ汁
おやつ：そふと

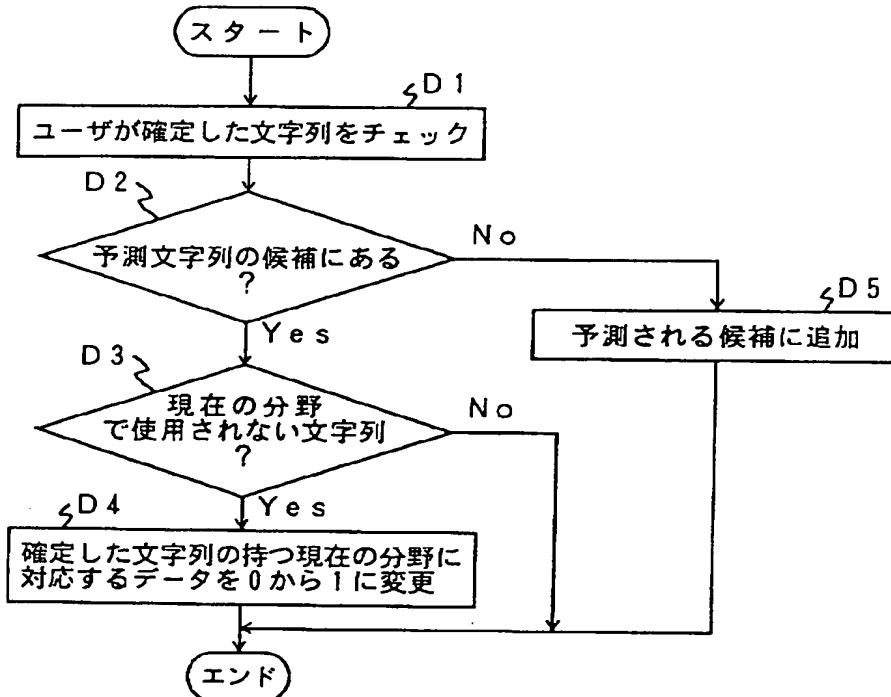
【図1-3】



【図1-9】



【図15】



【図17】

(a)

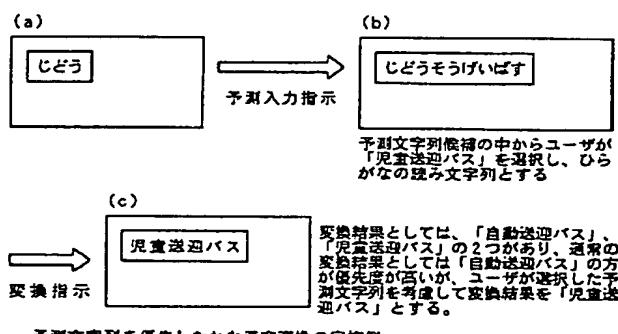
	食べ物	衣服	計算機	スポーツ	...	政治
ソフト	1	1	1	1		1
ソフトウェア	0	1	1	0		0
ソフトボール	0	0	0	1		0
ソフトクリーム	1	0	0	0		0
...						

(b)

	食べ物	衣服	計算機	スポーツ	...	政治
ソフト	1	1	1	1		1
ソフトウェア	0	1	1	0		0
ソフトボール	0	0	0	1		0
ソフトクリーム	1	0	1	0		0
...						

分野毎使用可能状況データの更新

【図22】



【図16】

(a)

パソコンの世界は競争が激しく、新しい製品が次々と作られ、発表されている。特に、ソフトウェアの世界では、その傾向が著しく毎月のように新しいソフトウェアが製品化されている。今月製品化された製品の中で特に注目を浴びているのが、「**そふと**

ソフトウェア

(b)

コンピュータ関係の文書ということで「ソフトウェア」が予測されたがユーザが入力したい「ソフトクリーム」ではないので、そのまま続けてユーザが入力する。

(c)

「そふとくりーむ」というそふとうえあである。」という読み文字列を入力し、かな漢字変換を行い、文字列を確定する。

パソコンの世界は競争が激しく、新しい製品が次々と作られ、発表されている。特に、ソフトウェアの世界では、その傾向が著しく毎月のように新しいソフトウェアが製品化されている。今月製品化された製品の中で特に注目を浴びているのが、「**ソフトクリーム**」というソフトウェアである。

確定後、文字列「ソフトクリーム」をコンピュータ分野の文書作成中にも予測できるように予測入力用辞書を変更する。

予測された文字列以外の文字列を入力、確定した場合の実施例

【図18】

(a)

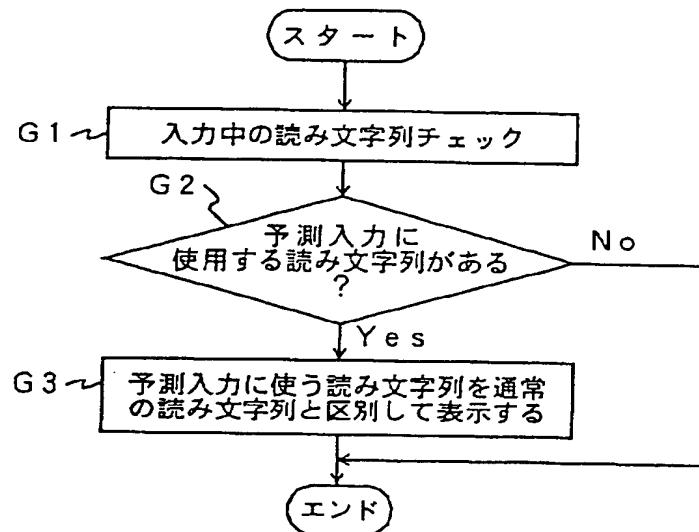
	食べ物	衣服	計算機	スポーツ	...	政治
ソフト	1	1	1	1		1
ソフトウェア	0	1	1	0		0
ソフトボール	0	0	0	1		0
ソフトクリーム	1	0	0	0		0
ソフトドリンク	1	0	0	1		0
ソフトカラー	0	1	0	0		0
ソフトフォーカス	0	0	0	0		0
...						

(b)

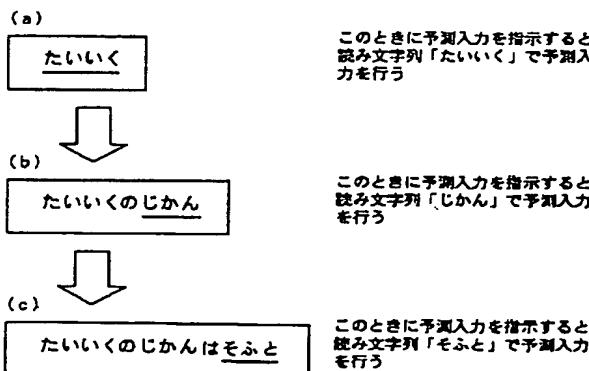
	食べ物	衣服	計算機	スポーツ	...	政治
ソフト	1	1	1	1		1
ソフトウェア	0	1	1	0		0
ソフトボール	0	0	0	1		0
ソフトクリーム	1	0	0	0		0
ソフトドリンク	1	0	0	1		0
ソフトカラー	0	1	0	0		0
ソフトフォーカス	0	0	0	0		0
ソフトハウス	0	0	1	0		0
...						

分野毎使用可能状況データに新しい予測文字列を追加

【図23】

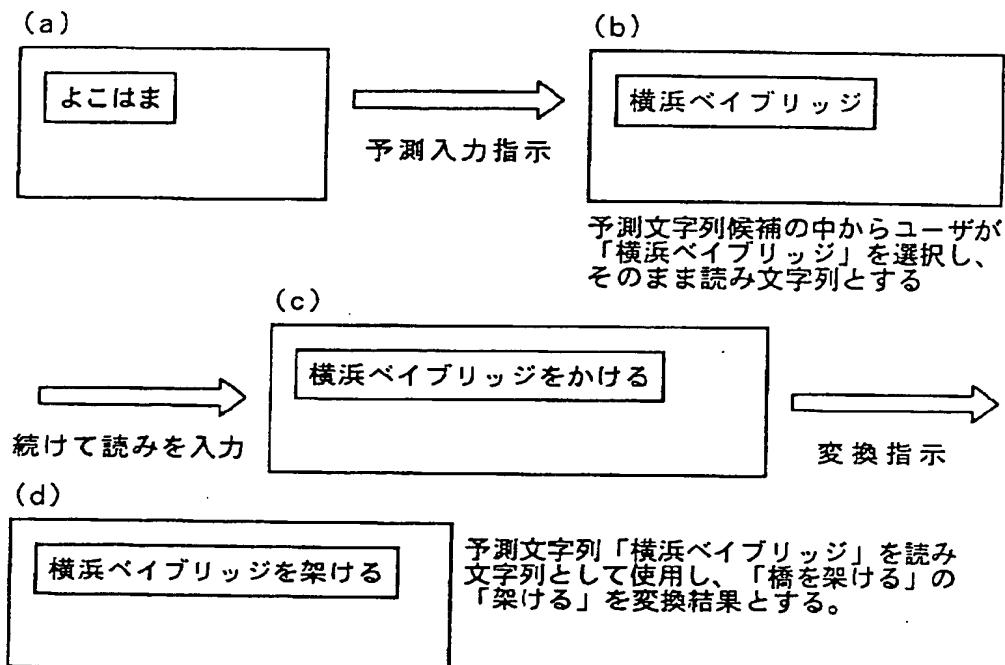


【図24】



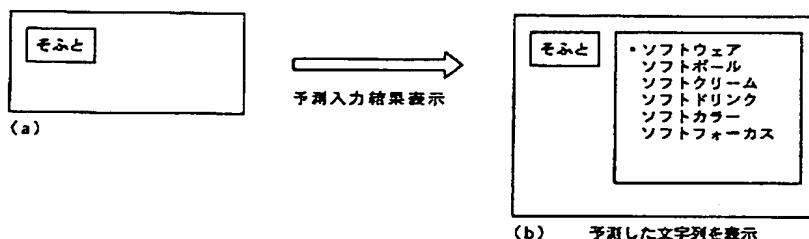
予測入力に使用する読み文字列の表示例

【図20】

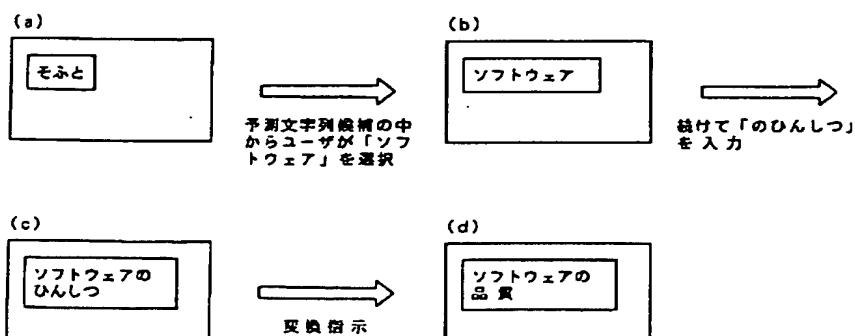


予測文字列をそのままかな漢字変換の読み文字列とするかな漢字変換の実施例

【図25】



【図26】



【図21】

